

PATENT APPLICATION
NUCLEIC ACIDS THAT CONTROL PLANT DEVELOPMENT

Inventor(s):

Robert L. Fischer, a citizen of the United States
residing at: 1423 Scott Street, El Cerrito, CA 94530

Yeonhee Choi, a citizen of South Korea
residing at: 4 Anchor Dr. #442, Emeryville, CA 94608

Mike Hannon, a citizen of the United States
residing at: 309 Pearl Drive, Livermore, CA 94550

Jack Kishiro Okamuro, a citizen of the United States
residing at: 307 Ocho Rios Way, Oak Park, CA 91377

Tatiana Valerievna Tatarinova, a citizen of Russia
residing at 836 North Beaudry Avenue Apt. 2, Los Angeles, CA 90012

Assignee:

THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
1111 Franklin Street, Fifth Floor
Oakland, CA 94607-5200

Entity:

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 415-576-0200

1020420715-099910US

TOE200-E4/04860

NUCLEIC ACIDS THAT CONTROL PLANT DEVELOPMENT

CROSS-REFERENCES TO RELATED APPLICATIONS

5 This application is a continuation-in-part of U.S. Patent Application No. 09/553,690, filed April 21, 2000, the contents of which are incorporated by reference.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

10 This invention was made with Government support under Grant No. 97-35304-4941, awarded by the United States Department of Agriculture. The government has certain rights in this invention.

FIELD OF THE INVENTION

This invention is directed to plant genetic engineering. It relates to, for 15 example, modulating seed (and in particular endosperm, embryo and seed coat) development, flowering time, chromosomal DNA methylation and modulating transcription in plants.

BACKGROUND OF THE INVENTION

20 A fundamental problem in biology is to understand how seed development. In flowering plants, the ovule generates the female gametophyte, which is composed of egg, central, synergid and antipodal cells (Reiser, *et al.*, *Plant Cell*, 1291-1301 (1993)). All are haploid except the central cell which contains two daughter nuclei that fuse prior to fertilization. One sperm nucleus fertilizes the egg to form the zygote, 25 whereas another sperm nucleus fuses with the diploid central cell nucleus to form the triploid endosperm nucleus (van Went, *et al.*, *Embryology of Angiosperms*, pp. 273-318 (1984)). The two fertilization products undergo distinct patterns of development. In *Arabidopsis*, the embryo passes through a series of stages that have been defined morphologically as preglobular, globular, heart, cotyledon and maturation (Goldberg, R. 30 B., *et al.*, *Science* (1994) 266: 605-614; Mansfield, S. G., *et al.*, *Arabidopsis: An Atlas of Morphology and Development*, pp. 367-383 (1994)). The primary endosperm nucleus undergoes a series of mitotic divisions to produce nuclei that migrate into the expanding central cell (Mansfield, S. G., *et al.*, *Arab Inf Serv* 27: 53-64 (1990); Webb, M. C., *et al.*, *Planta* 184:187-195 (1991)). Cytokinesis sequesters endosperm cytoplasm and nuclei

into discrete cells (Mansfield, S. G., *et al.*, *Arab Inf Serv* 27:65-72 (1990)) that produce storage proteins, starch, and lipids which support embryo growth (Lopes, M. A. *et al.*, *Plant Cell* 5:1383-1399 (1993)). Fertilization also activates development of the integument cell layers of the ovule that become the seed coat, and induces the ovary to grow and form the fruit, or silique, in *Arabidopsis*.

Of particular interest are recent discoveries of genes that control seed, and in particular endosperm, development. For instance, *MEDEA* (*MEA*) (also known as *FIE1* (see, e.g., copending U.S. patent application 09/071,838) and F644 (see, e.g., Kiyosue T, *et al.* (1999) *Proc Natl Acad Sci U S A* 96(7):4186-91) encodes an

- 10 *Arabidopsis* SET domain polycomb protein that appears to play a role in endosperm development. Inheritance of a maternal loss-of-function *mea* allele results in embryo abortion and prolonged endosperm production, irrespective of the genotype of the paternal allele. Thus, only the maternal wild-type *MEA* allele is required for proper embryo, endosperm, and seed coat development (Kinoshita T, *et al.* (1999) *Plant Cell* 10:1945-52). These results reveal functions for plant polycomb proteins in the suppression of central cell proliferation and endosperm development (Kiyosue T, *et al. supra*).
- 15

- 20 Another gene product that controls seed development is *FIE*, also known as *FIE3* (see, e.g., copending U.S. patent application 09/071,838). The *FIE* protein is a homolog of the WD motif-containing Polycomb proteins from *Drosophila* and mammals (Ohad, N. *et al.* *Plant Cell* 11(3):407-16 (1999)). In *Drosophila*, these proteins function as repressors of homeotic genes. Loss of function mutations in the *FIE* gene result in endosperm phenotypes that are identical to *medea* loss of function mutations. A female gametophyte with a loss-of-function allele of *fie* undergoes replication of the central cell nucleus and initiates endosperm development without fertilization. These results suggest that the *FIE* Polycomb protein functions to suppress a critical aspect of early plant reproduction, namely, endosperm development, until fertilization occurs. Moreover, hypomethylation of *fie* mutants leads to the development of differentiated endosperm. Vinkenoog *et al.*, *Plant Cell* 12:2271-2282 (2000).
- 25

- 30 Control of the expression of genes that control egg and central cell differentiation, or those that control reproductive development, i.e. embryo, endosperm and seed coat, is useful in the production of plants with a range of desired traits. These and other advantages are provided by the present application.

SUMMARY OF THE INVENTION

This invention provides isolated nucleic acids comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide comprising an amino acid sequence with at least 70% sequence identity to at least one of the following consensus sequences:

DMT Domain A

KV<1>(I, 1)D(D, p)(E, v)T<3>W<1>(L, v)L(M, 1)(E, d)<0-
2>D(K, e)<1>(K, t)<1>(K, a)(W, k)(W, l)<1>(E, k)ER<2>F<1>(G, t)R<1>(D, n)(S, 1)FI(A, n)RM(H, r)<1>(V, l)QG(D, n)R<1>F<1>(P, q)WKGSVVDSV(I, v)GVFLTQN(V, t)D(H, y)(L, s)SS(S, n)A(F, 10)yM<1>(L, v)A(A, s)<1>FF

DMT Domain B

W(D, n)<1>(L, f)R<5>E<3-
6>D(S, t)<1>(D, n)(Y, w)<3>R<10>I<2>RG(M, q)(N, f)<2>L(A, s)<1>RI<2-
12>FL<3>V<2>(H, n)G<1>IDLEWLRL<2>(P, d)(P, s)(D, h)<1>(A, v)K<1>(Y, f)LL(S, e)(I, f)<1>G(L, i)GLKS(V, a)ECVRLLL<1>L(H, k)<2>AFPVDTNVGR(I, a)cVR(M, 1)G(W, 1)VPL(Q, e)PLP<2>(L, v)Q(L, m)H(L, q)L(E, f)<1>YPL<1>(L, m)(E, d)(S, n)(I, v)QK(F, y)LWPRLLCKL(D, p)Q<1>TLIELHY(Q, h)(L, m)ITFGK<0>2>FCTK<2>PNCNACPM(R, k)<0>2>EC(R, k)(H, y)(F, y)(A, s)SA<1>(A, v)<0-10>S(A, s)(R, k)<1>(A, 1)L(P, e)<1>(P, t)

DMT Domain C.

P(I, 1)(I, v)E(E, f)P<1>(S, t)P<2>5>E<0>15>(D, a)IE(D, e)<4-
23>(I, v)P<1>I<1>(L, f)(N, d)<8>17>(S, a)<1>(A, d)LV<8>(I, 1)P<2-
5>(K, r)(L, m)K<4>LRTEH<1>V(Y, f)(E, v)LPD<1>H<1>(L, i)L(E, k)<1>(D, e)D(P, i)<2>YLL(A, s)25)IW(T, q)P(G, d)(E, g)<6>8>(P, s)<3>C<6-
10>(M, 1)C<4>C<2>C<3>(R, k)E<5>(V, f)RGT(L, i)L<0-
22>(L, v)FADH<1>(S, t)(S, r)<2>PI<3>(R, t)<3>(W, k)<1>L<1>(R, k)R<4>G(T, s)(S, t)<2>(S, t)I(F, c)(R, k)(G, 1)L<1>(T, v)<2>I<2>(C, n)F(W, q)<1>G(F, y)(V, 1)C(V, 1)R<1>F(E, d)<3>(R, g)<1>P(R, k)<1>L<2>(R, h)LH<2>(A, v)SK

In some embodiments, the nucleic acids of the invention do not encode a polypeptide at least 40% identical to SEQ ID NO:2, or alternatively at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% to SEQ ID NO:2. In some embodiments, the DMT polypeptide comprises an amino acid sequence 100% identical to the above-listed consensus sequences.

In some embodiments, the DMT polypeptides ar at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% identical to DMT domains A, B and/or C.

In one aspect, the invention provides DMt polypeptides capable of exhibiting at least one of the following biological activities:

- (a) glycosylase activity;
 - (b) demethylation of polynucleotides;
 - 5 (c) DNA repair;
 - (d) wherein expression of the polypeptide in a plant modulates organ identity;
 - (e) wherein expression of the polypeptide in a plant modulates organ number;
 - 10 (f) wherein expression of the polypeptide in a plant modulates meristem stem and/or activity;
 - (g) wherein enhanced expression of the polypeptide in a plant results in a delay in flowering time;
 - (h) wherein introduction of the polypeptide into a cell results in
 - 15 modulation of methylation of chromosomal DNA in the cell;
 - (i) wherein reduction of expression of the polypeptide in a plant results in modulation of endosperm development;
 - (j) wherein expression of the polypeptide in an *Arabidopsis* leaf results in modulation of expression of the MEDEA gene.
 - 20 In some aspects, the polypeptide comprises either a
 - (i) basic region;
 - (ii) nuclear localization signal;
 - (iii) leucine zipper;
 - (iv) helix-hairpin-helix structure;
 - (v) glycine-proline rich loop with a terminal aspartic acid or
 - (vi) helix that is capable of binding DNA.
- In one aspect, the invention provides methods of modulating in a plant one or more of the following:
- (a) DNA repair;
 - 25 (b) wherein expression of the polypeptide in a plant modulates organ identity;
 - (c) wherein expression of the polypeptide in a plant modulates organ number;

- TOC2010-ETH/00860
- (d) wherein expression of the polypeptide in a plant modulates meristem stem and/or activity;
- (e) wherein enhanced expression of the polypeptide in a plant results in a delay in flowering time;
- 5 (f) wherein introduction of the polypeptide into a cell results in modulation of methylation of chromosomal DNA in the cell;
- (g) wherein reduction of expression of the polypeptide in a plant results in modulation of endosperm development;
- (h) wherein expression of the polypeptide in an *Arabidopsis* leaf results in expression of the *MEDEA* gene,
- 10 wherein the method comprises:
- (a) introducing into a plant cell a nucleic acid of claim 1; and
- (b) generating conditions where the plant cell can transcribe the nucleic acid described above.
- 15 In some embodiments, the polypeptides comprise between 1500 and 2000 amino acids. In some aspects, the polypeptide has glycosylase activity. In some embodiments, introduction of the nucleic acid into a cell results in modulation of methylation of chromosomal DNA in the cell. In some embodiments, enhanced expression of the nucleic acids of the invention into a plant results in a delay in flowering time. In some embodiments, reduction of expression of a DMT polypeptide in a plant results in enhanced endosperm development. In addition, in some embodiments, expression of the nucleic acid of the invention in an *Arabidopsis* leaf results in expression of the *MEDEA* gene.
- 20 This invention provides isolated nucleic acids comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C. For instance, the nucleic acid can encode the DMT polypeptide displayed in SEQ ID NO:2. In one aspect, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide. In some embodiments, the promoter is constitutive. In other embodiments, the promoter is from a DMT gene. For example, the promoter can comprise a polynucleotide at least 70% identical to SEQ ID NO:3. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects of this invention, the promoter further comprises a polynucleotide at least

70% identical to SEQ ID NO:4. For example, in some aspects the promoter comprises SEQ ID NO:4. In some aspects, the polynucleotide sequence is linked to the promoter in an antisense orientation.

5 The invention also provides an isolated nucleic acid molecule comprising a polynucleotide sequence exhibiting at least 60% sequence identity to SEQ ID NO:1.

The invention also provides an expression cassette comprising a promoter operably linked to a heterologous polynucleotide sequence, or complement thereof, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2.

For instance, the nucleic acid can encode the DMT polypeptide displayed in SEQ ID

10 NO:2. In some aspects, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide. In some embodiments, the promoter is

constitutive. In other embodiments, the promoter is from a DMT gene. For example, the promoter can comprise a polynucleotide at least 70% identical to SEQ ID NO:3. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects of this invention, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For example, in some aspects the promoter comprises SEQ ID NO:4. In some aspects, the polynucleotide sequence is linked to the promoter in an antisense orientation.

15 The invention also provides an expression cassette for the expression of a heterologous polynucleotide in a plant cell. In some aspects, the expression cassette comprises a promoter polynucleotide at least 70% identical to SEQ ID NO:3 that is operably linked to a heterologous polynucleotide. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For instance, in some 20 embodiments, the promoter comprises SEQ ID NO:4. In some aspects, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:6. In some aspects, the promoter comprises SEQ ID NO:6.

25 The present invention also provides a host cell comprising an exogenous polynucleotide sequence comprising a polynucleotide sequence, or complement thereof, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide sequence. In some aspects, the promoter is constitutive. In some aspects, the promoter comprises a polynucleotide at least 70% identical to SEQ ID

NO:3. The promoter, for instance, can comprise SEQ ID NO:3. In some aspects, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For instance, in some embodiments, the promoter comprises SEQ ID NO:4. In some aspects, the promoter is operably linked to the exogenous polynucleotide sequence in an antisense orientation.

5

The present invention also provides an isolated polypeptide comprising an amino acid sequence at least 60% identical to SEQ ID NO:2 or an amino acid sequence at least 70% sequence identical to at least one of DMT domain A, B, or C and capable of exhibiting at least one biological activity of the polypeptide displayed in SEQ ID NO:2, or fragment thereof. The present invention also provides for an antibody capable of binding such polypeptides.

10

The present invention also provides a method of introducing an isolated nucleic acid into a host cell comprising, (a) providing an isolated nucleic acid or its complement, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C and (b) contacting the nucleic acid with the host cell under conditions that permit insertion of the nucleic acid into the host cell.

15

The present invention also provides a method of modulating transcription, comprising introducing into a host cell an expression cassette comprising a promoter operably linked to a heterologous DMT polynucleotide, the heterologous DMT polynucleotide encoding a DMT polypeptide at least 60% identical to SEQ ID NO:2 or at least 70% sequence identical to at least one of DMT domain A, B, or C, and detecting a host cell with modulated transcription. In some aspects of the invention, the heterologous DMT polynucleotide encodes SEQ ID NO:2. In some aspect, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects, the expression cassette is introduced into a host cell by *Agrobacterium*. In some aspects, the expression cassette is introduced by a sexual cross. In some aspects of the method of the invention, modulating transcription results in the modulation of endosperm development in a plant. In some aspects, endosperm development is enhanced. In other aspects, endosperm development is decreased. In some aspects of the methods of the invention, the promoter is operably linked to the DMT polynucleotide in an antisense orientation.

20

25

The present invention also provides a method of detecting a nucleic acid in a sample, comprising (a) providing an isolated nucleic acid molecule comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide exhibiting at

least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C, (b) contacting the isolated nucleic acid molecule with a sample under conditions that permit a comparison of the sequence of the isolated nucleic acid molecule with the sequence of DNA in the sample, and (c) analyzing the result of the comparison. In some aspects of the method, the isolated nucleic acid molecule and the sample are contacted under conditions that permit the formation of a duplex between complementary nucleic acid sequences.

The present invention also provides a transgenic plant cell or transgenic plant comprising a polynucleotide sequence, or its complement, encoding a DMT polypeptide exhibiting at least 60% sequence identity to SEQ ID NO:2 or exhibiting at least 70% sequence identity to at least one of DMT domain A, B, or C. For instance, the nucleic acid can encode the DMT polypeptide displayed in SEQ ID NO:2. In one aspect, the polynucleotide sequence comprises SEQ ID NO:5 or SEQ ID NO:1. In some aspects of the invention, the nucleic acid further comprises a promoter operably linked to the polynucleotide. In some embodiments, the promoter is constitutive. In other embodiments, the promoter comprises a polynucleotide at least 70% identical to SEQ ID NO:3. In some aspects, the promoter comprises SEQ ID NO:3. In some aspects of this invention, the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:4. For example, in some aspects the promoter comprises SEQ ID NO:4. In some aspects, the polynucleotide sequence is linked to the promoter in an antisense orientation. The present invention also provides a plant that is regenerated from a plant cell as described above.

The present invention also provides an expression cassette for the expression of a heterologous polynucleotide in a plant cell, wherein the expression cassette comprises a promoter at least 70% identical to SEQ ID NO:3 and the promoter is operably linked to a heterologous polynucleotide. In some embodiments, the promoter comprises a polynucleotide at least 70% identical to SEQ ID NO:4 and/or SEQ ID NO:6. In some embodiments, the promoter specifically directs expression of the heterologous polynucleotide in a female gametophyte when the expression cassette is introduced into a plant.

DEFINITIONS

The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It

includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role.

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants. Promoters include necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The term "plant" includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of flowering plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant, or a predecessor generation of the plant, by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include Agrobacterium-mediated transformation, biostatic methods, electroporation, in planta techniques, and the like. "Exogenous," as referred to within, is any polynucleotide, polypeptide or protein sequence, whether chimeric or not, that is initially or subsequently introduced into the genome of an individual host cell or the

PCT/US2010/04960

organism regenerated from said host cell by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include Agrobacterium-mediated transformation (of dicots - e.g. Salomon et al. EMBO J. 3:141 (1984); Herrera-Estrella et al. EMBO J. 2:987 (1983); of monocots, representative papers are those by Escudero et al., Plant J. 10:355 (1996), Ishida et al., Nature Biotechnology 14:745 (1996), May et al., Bio/Technology 13:486 (1995)), biolistic methods (Armaleo et al., Current Genetics 17:97 1990)), electroporation, in planta techniques, and the like.

Such a plant containing the exogenous nucleic acid is referred to here as a T0 for the primary transgenic plant and T1 for the first generation. The term "exogenous" as used herein is also intended to encompass inserting a naturally found element into a non-naturally found location.

The phrase "host cell" refers to a cell from any organism. Preferred host cells are derived from plants, bacteria, yeast, fungi, insects or other animals, including humans. Methods for introducing polynucleotide sequences into various types of host cells are well known in the art.

The "biological activity of a polypeptide" refers to any molecular activity or phenotype that is caused by the polypeptide. For example, the ability to transfer a phosphate to a substrate or the ability to bind a specific DNA sequence is a biological activity. One biological activity of DMT is glycosylase activity, i.e., cleavage of the nucleotide base from the nucleotide sugar). Another biological activity of DMT is to demethylate nucleotides (e.g., DMT has 5'-methylcytosine glycosylase activity). In addition, DMT has the ability to modulate endosperm production, as described herein, and to modulate flowering time in plants. For example, when DMT expression or DMT activity is increased in a plant, the flowering time of the plant is delayed. Moreover, expression of a DMT polypeptide in a plant tissue (e.g., a leaf) that does not typically express the *MEDEA* gene (Grossniklaus U, et al., *Science* 280(5362):446-50 (1998)) results in the expression of *MEDEA*.

Additional biological activities of DMT polypeptides include: nuclear localization (e.g., as localized by amino acids 43-78 of SEQ ID NO:2); the ability to modulate plant organ size and/or number; the ability to modulate meristem size and/or activity; and to perform DNA repair, including nucleotide methylation or demethylation and/or repair and/or removal of mis-matched nucleotides from DNA.

An "expression cassette" refers to a nucleic acid construct, which when introduced into a host cell, results in transcription and/or translation of an RNA or

polypeptide, respectively. Antisense or sense constructs that are not or cannot be translated are expressly included by this definition.

A "DMT nucleic acid" or "DMT polynucleotide sequence" of the invention is a subsequence or full length polynucleotide sequence of a gene which encodes a polypeptide involved in control of reproductive development and which, when the maternal allele is mutated or when DMT activity is reduced or eliminated in a maternal tissue or plant, allows for increased production of the endosperm and/or abortion of the embryo. In addition, overexpression of DMT in plants results in delayed time to flowering. Moreover, DMT is necessary and sufficient for expression of *MEDEA* in a plant cell. An exemplary nucleic acid of the invention is the *Arabidopsis DMT* sequence (SEQ ID NO:1). Additional *DMT* nucleic acid sequences from a variety of plant species are also provided (e.g., SEQ ID NOS: 7-70). *DMT* polynucleotides are defined by their ability to hybridize under defined conditions to the exemplified nucleic acids or PCR products derived from them. A *DMT* polynucleotide is typically at least about 30-40 nucleotides to about 7000, usually less than about 10,000 nucleotides in length. More preferably, *DMT* polynucleotides contain a coding sequence of from about 100 to about 5500 nucleotides, often from about 500 to about 3600 nucleotides in length. A *DMT* polypeptide is typically at least 500 amino acids, typically at least 1000 amino acids, more typically at least 1500 amino acids. In some embodiments, a *DMT* polypeptide comprises fewer than 2000 amino acids, more typically fewer than 3000 amino acid and still more typically fewer than 5000 or 7500 amino acid in length.

As described below, *DMT* nucleic acid sequences encode polypeptides with substantial identity to at least one of following the consensus sequences:

DMT Domain A

KV<1>(I,1)D(D,p)(E,v)T<3>W<1>(L,v)L(M,1)(E,d)<0-
2>D(K,e)<1>(K,t)<1>(K,a)(W,k)(W,l)<1>(E,k)ER<2>F<1>(G,t)R<1>(D,n)(S,1)FI(A,n)RM(H,r)<1>(V,1)QG(D,n)R<1>P<1>(P,q)WKGSVVDSV(I,v)GVFLTQN(V,t)D(H,y)(L,s)SS(S,n)A(F,y)M<1>(L,v)A(A,s)<1>FP

DMT Domain B

W(D,n)<1>(L,f)R<5>E<3->
6>D(S,t)<1>(D,n)(Y,w)<3>R<1>I<2>RG(M,q)(N,f)<2>L(A,s)<1>RI<2-
12>FL<3>V<2>(H,n)G<1>IDLEWL<2>(P,d)(P,s)(D,h)<1>(A,v)K<1>(Y,f)LL(S,e)(I,f)<1>G(L,i)GLKS(V,a)ECVRLL<1>L(H,k)<2>APFVDTNVGRI(A,c)VR(M,1)G(W,1)VPL(Q,e)PLP<2>(L,v)Q
35 (L,m)H(L,q)L(E,f)<1>YP<1>(L,m)(E,d)(S,n)(I,v)QK(F,y)LWPRLLCKL(D,p)Q<1>TLYELHY(Q,h)

```

) (L,m) ITFGK<0-2>FCTK<2>PNCNACPM(R,k) <0-2>EC(R,k) (H,y) (F,y) (A,s) SA<1>(A,v) <0-
10>S(A,s) (R,k)<1>(A,l)L(P,e)<1>(P,t)

DMT Domain C.
5 P(I,1) (I,v) E(E,f) P<1>(S,t) P<2-5>E<0-15>(D,a) IE(D,e) <4-
23>(I,v) P<1>I<1>(L,f) (N,d) <8-17>(S,a)<1>(A,d) LV<8>(I,1) P<2-
5>(K,r) (L,m) K<4>LRTEH<1>V(Y,f) (E,v) LPD<1>H<1>(L,i)L(E,k)<1>(D,e) D(P,i)<2>YLL(A,s
) IW(T,q) P(G,d) (E,g) <6-8>(P,s)<3>C<6-
10>(M,l) C<4>C<2>C<3>(R,k) E<5>(V,f) RGT(L,i)L<0-
10 22>(L,v) FADH<1>(S,t) (S,r) <2>PI<3>(R,t) <3>(W,k)<1>L<1>(R,k) R<4>G(T,s) (S,t) <2>(S,t
) I(F,c) (R,k) (G,1)L<1>(T,v) <2>I<2>(C,n) F(W,q)<1>G(F,y) (V,l) C(V,1) R<1>F(E,d)<3>(R,
g)<1>P(R,k)<1>L<2>(R,h) LH<2>(A,v) SK

```

In addition, the following consensus sequence spanning all three domains

15 were identified:

```

<9-14>(T,q) (A,i) (S,k) (I,1) <3>(A,r) (S,k) <1>(G,m) <2>
(S,r) (P,k) <2>(K,f) <2>(E,1) K<0-1>K<0-
3>(P,r) <2>(P,r) <1>(K,r) (K,r) (G,d) (R,k)<1>(G,v)<1>(K,g) <3-
5>(P,s) (P,k) <3>(S,n)<1>(I,1) <0-2>(Q,d)<9>(P,q)<4>(K,a) (P,s) <14-16>(P,a) <4>L<0-
10>D<1>(I,1) <0-4>(L,n) <12-46>(K,d) <2-
7>(P,a) KV<1>(I,1) D(D,p) (E,v) T<3>W<1>(L,v) L(M,1) (E,d) <0-
2>D(K,e) <1>(K,t) <1>(K,a) (W,k) (W,l) <1>(E,k) ER<2>F<1>(G,t) R<1>(D,n) (S,1) FI(A,n) RM(
H,r) <1>(V,1) QG(D,n) R<1><1>(P,q) WKGSVVDSV(I,v) GVFLLTON(V,t) D(H,y) (L,s) SS(S,n) A(F,
y) M<1>(L,v) A(A,s) <1>FP<0-16>(P,v) <6-15>(S,h) <3>(E,d) <10-
24>(N,t) <1>(S,e) <6>(N,k) <8-55>(E,i) <8-9>(I,v) <1>(N,s) <1-
4>(E,d) <1>(E,s) <4>(Q,1) <0-11>(D,h) <1>(P,m) <5>(Q,n) <0-3>(G,e) <2>(G,d) S<1>(K,d) <7-
11>(T,m) <2>(V,1) <3>(S,q) <6-10>(S,e) <2-3>(S,v) <19-25>(T,s) <16-28>(R,s) <2-
6>(T,p) <5>(P,k) <10>(Q,e) <4>(D,s) <1-
4>(S,r) <5>(D,p) <3>(N,d) <3>(P,y) <2>(F,s) <1>(R,k) <1>(G,s) <1>(S,a) (V,r) (P,e) <3>(T,s
30 ) <3-6>(I,1) <3>(P,e) <1>B<3-
5>(L,q) <1>(G,c) <1>(S,h) (S,n) <1>(V,q) <1>(E,d) <3>T(Q,e) <1-2>(N,g) <3>(E,n) <20-
30>(N,a) (P,g) <1-6>(S,1) <25-46>(Q,d) W(D,n) <1>(L,f) R<5>E<3-
6>D(S,t) <1>(D,n) (Y,w) <3>R<1>L<1>2>RG(M,q) (N,f) <2>L(A,s) <1>RI<2-
12>FL<3>V<2>(H,n) G<1>IDLEWLRL<2>(P,d) (P,s) (D,h) <1>(A,v) K<1>(Y,f) LL(S,e) (I,f) <1>G(
35 L,i) GLKS(V,a) ECVRLL<1>L(H,k) <2>APPVDTNVGRI(A,c) VR(M,1) G(W,1) VPL(Q,e) PLP<2>(L,v) Q
(L,m) H(L,q) L(E,f) <1>YP<1>(L,m) (E,d) (S,n) (I,v) QR(F,y) LWPRLCKL(D,p) Q<1>TLYELHY(Q,h
) (L,m) ITFGK<0-2>FCTK<2>PNCNACPM(R,k) <0-2>EC(R,k) (H,y) (F,y) (A,s) SA<1>(A,v) <0-
10>S(A,s) (R,k) <1>(A,l)L(P,e)<1>(P,t) (E,q) <7-16>P(I,1) (I,v) E(E,f) P<1>(S,t) P<2-
5>E<0-15>(D,a) IE(D,e) <4-23>(I,v) P<1>I<1>(L,f) (N,d) <8-
40 17>(S,s) <1>(A,d) LV<8>(I,1) P<2-
5>(K,r) (L,m) K<4>LRTEH<1>V(Y,f) (E,v) LPD<1>H<1>(L,i)L(E,k)<1>(D,e) D(P,i)<2>YLL(A,s
) IW(T,q) P(G,d) (E,g) <6-8>(P,s)<3>C<6-
10>(M,l) C<4>C<2>C<3>(R,k) E<5>(V,f) RGT(L,i)L<0-
22>(L,v) FADH<1>(S,t) (S,r) <2>PI<3>(R,t) <3>(W,k)<1>L<1>(R,k) R<4>G(T,s) (S,t) <2>(S,t

```

) I (F, c) (R, k) (G, 1) L<1>(T, v) <2>I<2>(C, n) F (W, q) <1>G (F, y) (V, 1) C (V, l) R<1>P (E, d) <3> (R, g) <1>P (R, k) <1>L<2>(R, h) L H<2>(A, v) SK

DMT domain A corresponds to amino acid positions 697 through 796 of SEQ ID NO:2. DMT domain B corresponds to amino acid positions 1192 through 1404 of SEQ ID NO:2. DMT domain C corresponds to amino acid positions 1452 through 1722 of SEQ ID NO:2. The consensus sequence provides amino acid sequences by position using single letter amino acid abbreviations. Numbers in carrots ("<" or ">") refer to amino acid positions where there is no consensus and which therefore, can be any amino acid. Amino acid abbreviations in parentheses indicate alternative amino acids at the same position. Capitalized letters refer to predominant consensus amino acids and lower case letters refer to amino acids that are commonly found in DMT sequences, but are not predominant. Thus, it is a simple matter to identify whether any particular nucleic acid sequence is a DMT nucleic acid and/or encodes a DMT polypeptide.

The structure of full-length DMT polypeptides comprises the following domains and regions. These regions are generally described with reference to SEQ ID NO:2. First, as described above, domain B DMT polypeptides can comprise a bipartite nuclear localization signal (e.g., amino acid positions 43-60 and 61-78 in SEQ ID NO:2) comprised of basic amino acids. Amino acids 36-91 are homologous to human G/T mismatch-specific thymine DNA glycosylase (Genbank accession number AAC50540.1), which has 5-methylcytosine glycosylase activity (Zhu *et al.*, *Nuc. Acids Res.* 28:4157-4165 (2000)). DMT polypeptides also contain a leucine zipper sequence (e.g., positions 1330-1351 of SEQ ID NO:2), that can be involved in protein-protein interactions as well as DNA binding. In addition, the amino portion of the DMT polypeptide (amino acids 43-78) is generally basic, similar to histone H1. Thus, without intending to limit the scope of the invention, it is believed this basic portion of DMT facilitates interactions with DNA and/or chromatic proteins.

In addition, amino acids 1-800 is related to the beta subunit of bacterial DNA-dependent RNA polymerases. Without intending to limit the scope of the invention, it is believed the RNA polymerase-like domain facilitates interaction of DMT with DNA.

Amino acids 1167-1368 is related to proteins in the HhH-GPD superfamily. Amino acids 1,271 to 1,304 correspond to the conserved HhH-GPD motif. The corresponding DMT sequence is

DKAKDYLLSIRGLGLKSVECVRLTLHNLAFPVD. Secondary structure prediction

(Jpred program) indicates that DMT has two alpha-helices (1,271 - 1,279 and 1,286 to 1,295) that correspond to the conserved alphaK and alphaL helices in the HhH-GPD motif of the crystallized hOGG1 DNA repair protein (Bruner et al *Nature* 403:859-866 (2000)). In between the two helices (1280 to 1285), is a hairpin with conserved glycines (G1282 and G1284). Amino acids 1286 to 1295 are related to the alphaL helix of hOGG1, which contacts the DNA backbone (Bruner et al *Nature* 403:859-866 (2000)). Thus, without intending to limit the scope of the invention, it is believed this region of DMT contacts the DNA. The catalytic lysine (K1286) and aspartic acid (D1304) residues are conserved in the HhH-GPD motif of DMT. Without intending to limit the scope of the invention, by analogy to hOGG1, K1286 is predicted to displace the modified base and to promote conjugate elimination of the 3'-phosphodiester bond. Without intending to limit the scope of the invention, by analogy to hOGG1, D1304 is believed to assist the reaction by transferring protons to and from K1286.

DMT nucleic acids are a new class of plant regulatory genes that encode polypeptides with sequence identity to members of the endonuclease III genes found in a diverse collection of organisms. Endonuclease III is implicated in various DNA repair reactions. Thus proteins related to endonuclease III are likely to have a chromosomal function. DMT (SEQ ID NO:1) is most related to endonuclease III from *Deinococcus radiodurans* Genbank Accession No. AE002073 (see, e.g., White, O. et al. *Science* 286:1571-1577 (1999)). DMT polypeptides have glycosylase activity (i.e., the capability to cleave the base portion of a nucleotide from the sugar portion). More particularly, DMT polypeptides have demethylase activity, and in more preferred embodiments, have 5-methylcytosine glycosylase activity. Demethylation activity can be assayed *in vivo* by expressing a candidate polypeptide in the nucleus of a cell and then assaying for a change in methylation of the cell's DNA. See, e.g., Vong, et al., *Science* 260:1926-1928 (1993). Changes in chromosomal methylation can be measured by comparing the ability of methylation sensitive and insensitive endonucleases to cleave DNA from a cell expressing a polypeptide suspected of having demethylase or methylase activity. Alternatively, bisulfate sequencing can be used to identify which base pairs are methylated in a DNA sequence. For a discussion of both methods, see Soppe et al., *Molec. Cell.* 6:791-802 (2000). *In vitro* assays to measure demethylase activity using labeled substrates are also known to those of skill in the art. See, e.g., Vhu et al., *Proc. Natl. Acad. Sci. USA* 97:5135-5139 (2000).

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only "substantially identical" to a sequence of the gene from which it was derived. As explained below, these substantially identical variants are specifically covered by the term *DMT nucleic acid*.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the terms "*DMT nucleic acid*". In addition, the term specifically includes those sequences substantially identical (determined as described below) with a DMT polynucleotide sequence disclosed here and that encode polypeptides that are either mutants of wild type DMT polypeptides or retain the function of the DMT polypeptide (e.g., resulting from conservative substitutions of amino acids in the DMT polypeptide). In addition, variants can be those that encode dominant negative mutants as described below.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a

score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, *Computer Applic. Biol. Sci.* 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

5 The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to a sequence or subsequence that has at least 40% sequence identity with a reference sequence. Alternatively, percent identity can be any integer from 40% to 100%. More preferred embodiments include at least: 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% compared to a reference sequence
10 using the programs described herein; preferably BLAST using standard parameters, as described below. This definition also refers to the complement of a test sequence, when the test sequence has substantial identity to a reference sequence.

15 For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program
20 parameters.

25 A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for
30 similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the

alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

5 The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5783-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance.

10 For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein.

20 For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

30 As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid.

Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 5 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 10 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(see, e.g., Creighton, *Proteins* (1984)).

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30°C below the T_m . The

T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 55°C, sometimes 60°C, and sometimes 65°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 time background hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such case, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

In the present invention, genomic DNA or cDNA comprising DMT nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C and sometimes 65°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

DETAILED DESCRIPTION

This invention provides molecular strategies for controlling plant development, including methylation of chromosomal DNA, endosperm development and flowering time.

5 Reproduction in flowering plants involves two fertilization events in the haploid female gametophyte. One sperm nucleus fertilizes the egg to form the embryo. A second sperm nucleus fertilizes the central cell to form the endosperm, a unique tissue that supports the growth of the embryo. Fertilization also activates maternal tissue differentiation, the ovule integuments form the seed coat and the ovary forms the fruit.

10 The present invention is based, at least in part, on the discovery of a set of female-gametophytic mutations and the subsequent cloning of the gene involved, termed *DEMETER (DMT)*, formerly known as *ATROPOS (ATR)*. Two mutant alleles of *DMT* disclosed here were created using a T-DNA tag, thereby disrupting an exon of the gene. The *dmt* mutations affect endosperm production, allowing for increased endosperm 15 development. Generally, the mutant *dmt* alleles are not transmitted by the female gametophyte. Inheritance of a mutant *dmt* allele by the female gametophyte usually results in embryo abortion and endosperm overproduction, even when the pollen bears the wild-type *DMT* allele.

In contrast, transmission of *dmt* mutant alleles through the male 20 gametophyte (*i.e.*, pollen) is ecotype-dependent in *Arabidopsis*. For instance, in some ecotypes (*e.g.*, Columbia), transmission of *dmt* mutant alleles is less than 50%. However, in Landsberg *erecta*, transmission is almost normal.

25 *DMT* is a repressor of endosperm both before and after fertilization. *DMT* is both necessary and sufficient for *MEDEA* transcription. *DMT* is related to 5-methylcytosine glycosylases. *DMT* regulates transcription of specific target genes (*i.e.*, MEA) by a demethylation mechanism. *DMT* is also required for maintaining the proper global pattern of methylation of chromosomal DNA in cells.

The isolated sequences prepared as described herein, can be used in a 30 number of techniques, for example, to suppress or enhance endogenous *DMT* gene expression. Modulation of *DMT* gene expression or *DMT* activity in plants is particularly useful, for example, in producing embryo-less or embryo-reduced seed, seed with increased endosperm, as part of a system to generate seed, to modulate time to flowering, organ identity, size and/or number, meristem size or activity in plants, or to modulate methylation, and thus gene expression in plants. Another use is the expression of *DMT*

polynucleotides in animal cells, for instance as a DNA repair enzyme useful in preventing the unnatural proliferation of cells (including cancer) due to chromosomal lesions. See, e.g., Bruner, *et al.*, *Nature* 403:859 (2000).

As described in more detail below, reduction of expression of DMT in plants results in a number of diverse phenotypes. Without intending to limit the invention to particular embodiments, it is believed that some of the phenotypes that are generated in plants are epigenetic mutations, i.e., effects due to differences in the methylation state of the chromosome that result in altered gene expression. Thus, DMT provides a powerful tool to develop any number of plant lines with a variety of desired phenotypes.

10

Isolation of DMT nucleic acids

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989).

20

The isolation of DMT nucleic acids may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired gene in a cDNA or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ, such as ovules, and a cDNA library which contains the DMT gene transcript is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which DMT genes or homologs are expressed.

25

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned DMT gene disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Alternatively, antibodies raised against a DMT polypeptide can be used to screen an mRNA expression library.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the *DMT* genes directly from genomic DNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. For a general overview of PCR see *PCR Protocols: A Guide to Methods and Applications*. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San Diego (1990).

Appropriate primers and probes for identifying *DMT* sequences from plant tissues are generated from comparisons of the sequences provided here with other related genes. For instance, *DMT* can be compared to the other endonuclease III genes, such as Genbank Accession No. AE002073. Using these techniques, one of skill can identify conserved regions in the nucleic acids disclosed here to prepare the appropriate primer and probe sequences. Primers that specifically hybridize to conserved regions in *DMT* genes can be used to amplify sequences from widely divergent plant species. Appropriate primers for amplification of the genomic region or cDNA of *DMT* include the following primers:

Xba-SKEN-7; CCTCTAGAGGAATTGTCGGCAAAATCGAG
SKB-8; GGAGAGACGGITATTGTCAACC
SKB-7; AAAAGTCTACAAGGGAGAGAGAGT
SKB-5; GTAGATGTACATACGTACC
SKEN-8; GCATCCTCCAACAAGTAACAATCCACTC
SKB-6; CACTGAGATTAATTCTCAGACTCG
SKEN-3.5; CTCAGGCAGTCATGCCGGAGAACAC
SKEN-3; CGAGGGCTGATCCGGGGATAGATATTIT
SKEN-2; CCCCGGATCAGCCCTCGAATTTC
SKEN-1; CCCCTGTCTACAAATTACCCACCTGG
SKEL-4; CTGACCCAAC TGCTTCTCTC
skes1.5; TCACCTGTTCTGAACAGACTGG
SKES-1.4; CAGCAGACGAGTCCATAATGCTCTGC
SKES-2.4; GGTTGCCTTCCACGACCACC
SKES-1; GGAAGGCCACGCAAAGCTGCAACTCAGG
SKES-2.45; GAGTTGCAGCTTGCCTGGCTTCC

SKES2.5; TTCAGACTCAGAGTCACCTTGC
SKES-2; ACCAGCAGCTTGTCTGGCC
SKES-3; CATGCCAGAGAAGCAGGGCTCC
SKES3.5; CGATGATACTGTCTTTCGAGC
SKES-6; CCTCCGCCTGCTCATGCCCTAG
SKEN-4; GTCCATCAGGAGAACCTCTGTGTCAGGAT
SKES-4; GGGAAACAAGTGCACCACATCTCC
SKEN-6; GCTCTCATAGGGAACAAGTGCACCACATCTC
SKES-5; CGCTCGCATGCACCTGGTAC
SKB-1; GGAGGGAATCGAGCAGCTAGAG
SKB-2; GAGCAGCTAAGGGACTGTTCAAACTC
SKB-3; CCAGGAATGGGATTGTCGG
3' RACE-2; CTTGGACGGCGCTTGAGGAACC
3' RACE-1; GCCTACAAGCCAGTGGGATAG
cDNA-1; GCCAAGGACTATCTTGTGAGC
SKB-4; GGATGGACTCGAGCACTGGG
SKE2.2-4; AGAGGAGAGTGCAGACACTTG
cDNA-3; GAGGACCTGACGAGATCCAAC
cDNA-9; CCATGTGTTCCCGTAGAGTCATTCC
2.2+SKE-1; ATGGAGCTCCAAGAAGGTGACATG
cDNA-5; CAGAACTGTGGAGGGAAAGCGTCTGGC
cDNA-4; CCCTCAGACTGTTACACTCAGAAC
cDNA-2; CCCGTGAGCGGAAAATTCCCTCATGGC
cDNA-7; GGAAAGGATTCGTATGTGTCCTGTGG
SKEN-5; GCAATGCGTTGCTTCTTCAGTCATCT
cDNA-6; GAGGAGAGCAGAGAAGCAATGCGTTGC
cDNA-8; GTTAGAGAGAAAATAATAACCC
2.2+SKE-3; CCGTAAACAAACACCGGATAACAC

The amplification conditions are typically as follows. Reaction components: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200 µM dATP, 200 µM dCTP, 200 µM dGTP, 200 µM dTTP, 0.4 µM primers, and 100 units per ml Taq polymerase. Program: 96 C for 3 min., 30 cycles of 96 C for 45 sec., 50 C for 60 sec., 72 for 60 sec, followed by 72 C for 5 min.

Standard nucleic acid hybridization techniques using the conditions disclosed above can then be used to identify full-length cDNA or genomic clones.

Alternatively, a number of methods for designing modifications of polynucleotide sequences are known to those of skill in the art. For example, oligonucleotide directed mutagenesis can be used to introduce site-specific mutations in a nucleic acid sequence of interest. Examples of such techniques are found in the references above and, e.g., in Reidhaar-Olson *et al.* *Science*, 241:53-57 (1988) and Ausubel *et al.* Similarly, gene shuffling (Stemmer *Proc. Natl. Acad. Sci. USA* 91:10747-10751(1994); Ostermeier *et al.* *Proc. Natl. Acad. Sci. USA*, 96: 3562-67(1999))) can be used to introduce variation into one or more DMT sequences or subsequences. For example, orthologous (between species) or homologous (within a species) DMT nucleic acids can be interchanged, combined or shuffled to produce novel variations within the scope of the invention.

Additionally, error prone PCR can also be used to introduce variation into a nucleic acid sequence. See, Leung *et al.* (1989) *Technique* 1:11-15 and Caldwell *et al.* (1992) *PCR Methods Applic.* 2:28-33.

15 *Control of DMT activity or gene expression*

Since DMT genes are involved in controlling seed, in particular endosperm, development, inhibition of endogenous DMT activity or gene expression is useful in a number of contexts. For instance, reduction of DMT activity can be used for production of seed with enhanced endosperm. By reducing and/or eliminating DMT activity, plants with seed containing increased endosperm can be produced.

Alternatively, substantial inhibition of DMT activity can be used for production of fruit with small and/or degraded seed (referred to here as "seedless fruit") after fertilization. In many plants, particularly dicots, the endosperm is not persistent and eventually is degraded. Thus, in plants of the invention in which DMT activity is inhibited, embryo-less seed do not persist and seedless fruit are produced. For production of dicots with enhanced endosperm, the most beneficial effect may be to reduce, but not eliminate DMT activity. On the other hand, in monocots, which have persistent endosperm, it is advantageous to eliminate DMT activity.

30 Alternatively, plants of the invention can be used to prevent pre-harvest sprouting in seeds, especially those derived from cereals. In these plants, the endosperm persists and is the major component of the mature seed. Premature growth of embryos in stored grain causes release of degradative enzymes which digest starch and other

components of the endosperm. Plants of the present invention are useful in addressing this problem because the seeds lack an embryo and thus will not germinate.

Moreover, as discussed herein, time to flowering and DNA methylation can also be modulated by modulating DMT activity in a cell. For example, DMT can be used to modulate the amount of methylated DNA in a cell. Indeed, since expression of many genes is dependent on their methylation state, modulation of DMT activity modulates gene expression in a cell. Examples of genes whose expression is modulated by DMT include *MEDEA*.

One of skill will recognize that a number of methods can be used to modulate DMT activity or gene expression. DMT activity can be modulated in the plant cell at the gene, transcriptional, posttranscriptional, translational, or posttranslational, levels. Techniques for modulating DMT activity at each of these levels are generally well known to one of skill and are discussed briefly below.

Methods for introducing genetic mutations into plant genes are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, for example, X-rays or gamma rays can be used.

Alternatively, homologous recombination can be used to induce targeted gene disruptions by specifically deleting or altering the *DMT* gene *in vivo* (see, generally, Grewal and Klar, *Genetics* 146: 1221-1238 (1997) and Xu *et al.*, *Genes Dev.* 10:2411-2422 (1996)). Homologous recombination has been demonstrated in plants (Puchta *et al.*, *Experientia* 50:277-284 (1994), Swoboda *et al.*, *EMBO J.* 13:484-489 (1994); Offringa *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 7346-7350 (1993); and Kempin *et al.* *Nature* 389:802-803 (1997)).

In applying homologous recombination technology to the genes of the invention, mutations in selected portions of a DMT gene sequences (including 5' upstream, 3' downstream, and intragenic regions) such as those disclosed here are made *in vitro* and then introduced into the desired plant using standard techniques. Since the efficiency of homologous recombination is known to be dependent on the vectors used, use of dicistronic gene targeting vectors as described by Mountford *et al.* *Proc. Natl. Acad. Sci. USA* 91:4303-4307 (1994); and Vaulont *et al.* *Transgenic Res.* 4:247-255 (1995) are conveniently used to increase the efficiency of selecting for altered DMT gene

expression in transgenic plants. The mutated gene will interact with the target wild-type gene in such a way that homologous recombination and targeted replacement of the wild-type gene will occur in transgenic plant cells, resulting in suppression of DMT activity.

Alternatively, oligonucleotides composed of a contiguous stretch of RNA and DNA residues in a duplex conformation with double hairpin caps on the ends can be used. The RNA/DNA sequence is designed to align with the sequence of the target *DMT* gene and to contain the desired nucleotide change. Introduction of the chimeric oligonucleotide on an extrachromosomal T-DNA plasmid results in efficient and specific DMT gene conversion directed by chimeric molecules in a small number of transformed plant cells. This method is described in Cole-Strauss *et al. Science* 273:1386-1389 (1996) and Yoon *et al. Proc. Natl. Acad. Sci. USA* 93:2071-2076 (1996).

Gene expression can be inactivated using recombinant DNA techniques by transforming plant cells with constructs comprising transposons or T-DNA sequences.

DMT mutants prepared by these methods are identified according to standard techniques.

For instance, mutants can be detected by PCR or by detecting the presence or absence of *DMT* mRNA, e.g., by Northern blots. Mutants can also be selected by assaying for development of endosperm in the absence of fertilization.

The isolated nucleic acid sequences prepared as described herein, can also be used in a number of techniques to control endogenous *DMT* gene expression at various levels. Subsequences from the sequences disclosed here can be used to control, transcription, RNA accumulation, translation, and the like.

A number of methods can be used to inhibit gene expression in plants. For instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense suppression can act at all levels of gene regulation including suppression of RNA translation (see, Bourque *Plant Sci. (Limerick)* 105:125-149 (1995); Pantopoulos In Progress in Nucleic Acid Research and Molecular Biology, Vol. 48. Cohn, W. E. and K. Moldave (Ed.). Academic Press, Inc.: San Diego, California, USA; London, England, UK. p. 181-238; Heiser *et al. Plant Sci. (Shannon)* 127:61-69 (1997)) and by preventing the accumulation of mRNA which encodes the protein of interest, (see, Baulcombe *Plant Mol. Bio.* 32:79-88 (1996); Prins and Goldbach *Arch. Virol.* 141:2259-2276 (1996);

Metzlaff *et al.* *Cell* 88:845-854 (1997), Sheehy *et al.*, *Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt *et al.*, U.S. Patent No. 4,801,340).

The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous *DMT* gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other genes within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA.

Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of about 500 to about 7000 nucleotides is especially preferred.

A number of gene regions can be targeted to suppress *DMT* gene expression. The targets can include, for instance, the coding regions, introns, sequences from exon/intron junctions, 5' or 3' untranslated regions, and the like. In some embodiments, the constructs can be designed to eliminate the ability of regulatory proteins to bind to *DMT* gene sequences that are required for its cell- and/or tissue-specific expression. Such transcriptional regulatory sequences can be located either 5'-, 3'-, or within the coding region of the gene and can be either promote (positive regulatory element) or repress (negative regulatory element) gene transcription. These sequences can be identified using standard deletion analysis, well known to those of skill in the art. Once the sequences are identified, an antisense construct targeting these sequences is introduced into plants to control gene transcription in particular tissue, for instance, in developing ovules and/or seed. In one embodiment, transgenic plants are selected for DMT activity that is reduced but not eliminated.

Oligonucleotide-based triple-helix formation can be used to disrupt DMT gene expression. Triplex DNA can inhibit DNA transcription and replication, generate site-specific mutations, cleave DNA, and induce homologous recombination (*see, e.g.*, Havre and Glazer *J. Virology* 67:7324-7331 (1993); Scanlon *et al.* *FASEB J.* 9:1288-1296 (1995); Giovannangeli *et al.* *Biochemistry* 35:10539-10548 (1996); Chan and Glazer *J.*

Mol. Medicine (Berlin) 75:267-282 (1997)). Triple helix DNAs can be used to target the same sequences identified for antisense regulation.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of *DMT* genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. Thus, ribozymes can be used to target the same sequences identified for antisense regulation.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Zhao and Pick *Nature* 365:448-451 (1993); Eastham and Ahlering *J. Urology* 156:1186-1188 (1996); Sokol and Murray *Transgenic Res.* 5:363-371 (1996); Sun *et al. Mol. Biotechnology* 7:241-251 (1997); and Haseloff *et al. Nature*, 334:585-591 (1988).

Another method of suppression is sense cosuppression. Introduction of nucleic acid configured in the sense orientation has been recently shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes (see, Assaad *et al. Plant Mol. Bio.* 22:1067-1085 (1993); Flavell *Proc. Natl. Acad. Sci. USA* 91:3490-3496 (1994); Stam *et al. Annals Bot.* 79:3-12 (1997); Napoli *et al., The Plant Cell* 2:279-289 (1990); and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184).

The suppressive effect may occur where the introduced sequence contains no coding sequence *per se*, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous

sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

5 For sense suppression, the introduced sequence, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants that are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence
10 need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used. In addition, the same gene regions noted for antisense regulation can be targeted using cosuppression technologies.

In a preferred embodiment, expression of a nucleic acid of interest can be
15 suppressed by the simultaneous expression of both sense and antisense constructs (Waterhouse *et al.*, *Proc. Natl. Acad. Sci. USA* 95:13959-13964 (1998). See also Tabara *et al.* *Science* 282:430-431 (1998).

20 Alternatively, *DMT* activity may be modulated by eliminating the proteins that are required for *DMT* cell-specific gene expression. Thus, expression of regulatory proteins and/or the sequences that control *DMT* gene expression can be modulated using the methods described here.

Another method is use of engineered tRNA suppression of *DMT* mRNA translation. This method involves the use of suppressor tRNAs to transactivate target genes containing premature stop codons (see, Betzner *et al.* *Plant J.* 11:587-595 (1997);
25 and Choisne *et al.* *Plant J.* 11:597-604 (1997). A plant line containing a constitutively expressed *DMT* gene that contains an amber stop codon is first created. Multiple lines of plants, each containing tRNA suppressor gene constructs under the direction of cell-type specific promoters are also generated. The tRNA gene construct is then crossed into the *DMT* line to activate *DMT* activity in a targeted manner. These tRNA suppressor lines
30 could also be used to target the expression of any type of gene to the same cell or tissue types.

DMT proteins may form homogeneous or heterologous complexes *in vivo*. Thus, production of dominant-negative forms of *DMT* polypeptides that are defective in their abilities to bind to other proteins in the complex is a convenient means to inhibit

endogenous DMT activity. This approach involves transformation of plants with constructs encoding mutant DMT polypeptides that form defective complexes and thereby prevent the complex from forming properly. The mutant polypeptide may vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain. Use of dominant negative mutants to inactivate target genes is described in Mizukami *et al.* *Plant Cell* 8:831-845 (1996).

Another strategy to affect the ability of a DMT protein to interact with itself or with other proteins involves the use of antibodies specific to DMT. In this method cell-specific expression of DMT-specific Abs is used inactivate functional domains through antibody:antigen recognition (*see*, Hupp *et al.* *Cell* 83:237-245 (1995)).

After plants with reduced DMT activity are identified, a recombinant construct capable of expressing low levels of DMT in embryos can be introduced using the methods discussed below. In this fashion, the level of DMT activity can be regulated to produce preferred plant phenotypes. For example, a relatively weak promoter such as the ubiquitin promoter (*see*, e.g., Garbarino *et al.* *Plant Physiol.* 109(4):1371-8 (1995); Christensen *et al.* *Transgenic Res.* 5(3):213-8 (1996); and Holtorf *et al.* *Plant. Mol. Biol.* 29(4):637-46 (1995)) is useful to produce plants with reduced levels of DMT activity or expression. Such plants are useful for producing, for instance, plants that produce seed with enhanced endosperm.

Use of nucleic acids of the invention to enhance DMT gene expression

Isolated sequences prepared as described herein can also be introduced into a plant cell, thereby modulating expression of a particular *DMT* nucleic acid to enhance or increase endogenous gene expression. For instance, without being bound to any theory, in light of DMT's relation to Exonuclease III and DNA glycosylases, applicants believe that DMT binds DNA or chromatin and acts to modulate transcription by modulating the methylation state of DNA. Enhanced expression can therefore be used to control plant morphology by controlling expression of genes under DMT's control, such as *MEDEA*, in desired tissues or cells. Enhanced expression can also be used, for instance, to increase vegetative growth by preventing the plant from setting seed. Where overexpression of a gene is desired, the desired gene from a different species may be used to decrease potential sense suppression effects.

Moreover, as discussed herein, time to flowering and DNA methylation can also be modulated by modulating DMT activity in a cell. For example, increased expression of DMT in a plant results in delayed time to flowering. Similarly, DMT can be used to modulate the amount of methylated DNA in a cell. Indeed, since expression of many genes is dependent on their methylation state, modulation of DMT activity modulates gene expression in a cell. Examples of genes whose expression is modulated by DMT include *MEDEA*.

One of skill will recognize that the polypeptides encoded by the genes of the invention, like other proteins, have different domains that perform different functions. Thus, the gene sequences need not be full length, so long as the desired functional domain of the protein is expressed.

Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described in detail, below. For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

Preparation of recombinant vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of flowering plant species are well known and described in the technical and scientific literature. See, for example, Weising *et al. Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for overexpression, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of

Agrobacterium tumefaciens, and other transcription initiation regions from various plant genes known to those of skill. Such genes include for example, *ACT11* from *Arabidopsis* (Huang *et al.* *Plant Mol. Biol.* 33:125-139 (1996)), *Cat3* from *Arabidopsis* (GenBank No. U43147, Zhong *et al.*, *Mol. Gen. Genet.* 251:196-203 (1996)), the gene encoding 5 stearoyl-acyl carrier protein desaturase from *Brassica napus* (Genbank No. X74782, Solcombe *et al.* *Plant Physiol.* 104:1167-1176 (1994)), *Gpc1* from maize (GenBank No. X15596, Martinez *et al.* *J. Mol. Biol.* 208:551-565 (1989)), and *Gpc2* from maize (GenBank No. U45855, Manjunath *et al.*, *Plant Mol. Biol.* 33:97-112 (1997)).

Alternatively, the plant promoter may direct expression of the *DMT* 10 nucleic acid in a specific tissue or may be otherwise under more precise environmental or developmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light. Such promoters are referred to here as "inducible" or "tissue-specific" promoters. One of skill will recognize that a tissue-specific promoter may drive 15 expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue, but may also lead to some expression in other tissues as well.

Examples of promoters under developmental control include promoters 20 that initiate transcription only (or primarily only) in certain tissues, such as fruit, seeds, or flowers. Promoters that direct expression of nucleic acids in ovules, flowers or seeds are particularly useful in the present invention. As used herein a seed-specific promoter is one which directs expression in seed tissues, such promoters may be, for example, ovule-specific (which includes promoters which direct expression in maternal tissues or the female gametophyte, such as egg cells or the central cell), embryo-specific, endosperm-specific, integument-specific, seed coat-specific, or some combination thereof. Examples 25 include a promoter from the ovule-specific *BEL1* gene described in Reiser *et al.* *Cell* 83:735-742 (1995) (GenBank No. U39944). Other suitable seed specific promoters are derived from the following genes: *MAC1* from maize (Sheridan *et al.* *Genetics* 142:1009-1020 (1996), *Cat3* from maize (GenBank No. L05934, Abler *et al.* *Plant Mol. Biol.* 22:10131-1038 (1993), the gene encoding oleosin 18kD from maize (GenBank No. J05212, Lee *et al.* *Plant Mol. Biol.* 26:1981-1987 (1994)), viviparous-1 from *Arabidopsis* (Genbank No. U93215), the gene encoding oleosin from *Arabidopsis* (Genbank No. Z17657), Atmycl from *Arabidopsis* (Urao *et al.* *Plant Mol. Biol.* 32:571-576 (1996), the 2s seed storage protein gene family from *Arabidopsis* (Conceicao *et al.* *Plant* 5:493-505

(1994)) the gene encoding oleosin 20kD from *Brassica napus* (GenBank No. M63985), *napA* from *Brassica napus* (GenBank No. J02798, Josefsson *et al.* *JBL* 26:12196-1301 (1987), the napin gene family from *Brassica napus* (Sjodahl *et al.* *Planta* 197:264-271 (1995), the gene encoding the 2S storage protein from *Brassica napus* (Dasgupta *et al.* *Gene* 133:301-302 (1993)), the genes encoding oleosin A (Genbank No. U09118) and oleosin B (Genbank No. U09119) from soybean and the gene encoding low molecular weight sulphur rich protein from soybean (Choi *et al.* *Mol Gen, Genet.* 246:266-268 (1995)).

In addition, the promoter sequences from the *DMT* genes disclosed here
10 can be used to drive expression of the *DMT* polynucleotides of the invention or heterologous sequences. The sequences of the promoters are identified below.

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

15 The vector comprising the sequences (e.g., promoters or coding regions) from genes of the invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or
20 Basta.

Promoter and Enhancer Nucleic Acids of the Invention

The present invention provides polynucleotides useful as promoters and enhancers. The invention also provides methods of targeting heterologous polypeptides to a female gametophyte of a plant, including, e.g., the polar nuclei, the eggs and synergids and central cells. Promoter polynucleotides of the invention include, for example, sequences and subsequences of the DMT 5' flanking DNA (SEQ ID NO:3), the 5' UTR region (SEQ ID NO:6) and the 3' flanking region (SEQ ID NO:4). In some embodiments, the promoter sequences are operably linked to the 5' end of the DMT coding region, which is in turn fused to a polynucleotide of interest, typically encoding a polypeptide. An exemplary promoter sequence includes the last 3424 nucleotides of SEQ ID NO:3 linked to the first 1478 nucleotides of SEQ ID NO:5. In some embodiments, a further 444 nucleotides (e.g., the first 444 nucleotides of the DMT coding region) are incorporated into the promoter. In some embodiments, the promoter

sequences of the invention specifically direct expression of polynucleotides to the female gametophyte and does not direct expression in tissues following fertilization.

Production of transgenic plants

5 DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as
10 DNA particle bombardment.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al. *Embo J.* 3:2717-2722 (1984).

Electroporation techniques are described in Fromm et al. *Proc. Natl. Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein et al. *Nature* 327:70-73 (1987).

Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch et al. *Science* 233:496-498 (1984), and Fraley et al. *Proc. Natl. Acad. Sci. USA* 80:4803 (1983).

25 Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype such as increased seed mass. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been
30 introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants,

organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pannesetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna*, and *Zea*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

Seed obtained from plants of the present invention can be analyzed according to well known procedures to identify plants with the desired trait. If antisense or other techniques are used to control DMT gene expression, Northern blot analysis can be used to screen for desired plants. In addition, the presence of fertilization independent reproductive development can be detected. Plants can be screened, for instance, for the ability to form embryo-less seed, form seed that abort after fertilization, or set fruit in the absence of fertilization. These procedures will depend, part on the particular plant species being used, but will be carried out according to methods well known to those of skill.

25 **DMT Mutations, Fragments And Fusions**

As discussed above, DMT polynucleotides and polypeptides are not limited to the sequences disclosed herein. Those of skill in the art that conservative amino acid substitutions, as well as amino acid additions or deletions may not result in any change in biological activity. Moreover, sequence variants with at least one modulated biological activity of DMT are also contemplated. For example, at least one DMT activity can be increased or decreased by introduction of single or multiple amino acid changes from the sequences disclosed herein. Those of skill in the art will recognize that conservative amino acid substitutions in important functional domains are typically

useful in generating more active DMT polypeptides. Conversely, non-conservative substitutions of amino acid residues in functional domains, such as the HhH region of DMT (e.g., amino acids 1271-1304 of SEQ ID NO:2) are likely to disrupt at least one biological activity such as DNA binding. In some embodiments, the fragments of the invention consist of about 100, 200, 300 400, 500, 600, 700, 800, 900, or 1000 amino acids.

Alternatively, fragments of the sequences disclosed herein are contemplated. In some preferred embodiments, the polypeptide fragments have at least one biological activity of DMT. For example, amino acid sequences comprising DMT domain B represent polypeptide fragments with glycosylase or demethylase activity. In some embodiments, a fragment comprising amino acids 1167-1404, 1192-1404, 1192-1368 or 1167-1368 of SEQ ID NO:2 have glycosylase activity.

Mutations, fragments and fusions are also useful as dominant negative mutations. For instance, different regions of the DMT protein are responsible for different biological activities. Thus, mutation or deletion of one functional domain can eliminate one but not all activities. For example, mutation or deletion of the DNA binding domain may result in proteins that interact with proteins necessary for DMT function, effectively titrating out those proteins and preventing an active DMT protein from acting. Similarly, DMT fragments comprising the DNA binding portion of the protein with an inactive enzymatic domain or lacking an enzymatic domain are also useful as dominant negative mutants by competing with active DMT polypeptides for DNA binding sites. As described herein, domains of DMT that can be modulated include: the leucine zipper, nuclear localization sequence, HhH domain, the aspartic acid of the GPD domain, as well a DMT domains A, B or C. Without intending to limit the scope of the invention, based on the data provided herein, DMT has glycosylase and demethylase activity and is a DNA repair enzyme.

Targeting the polypeptides of the invention to chromosomal regions

Without intending to limit the scope of the invention, based on the data provided herein, it is believed that DMT has glycosylase and/or demethylase activity and is a DNA repair enzyme. DNA methylation plays an important role in the repression of gene transcription during animal development including embryogenesis, myogenesis and blood cell development. Methylated DNA is recognized by MeCP2 which in turn represses gene transcription by recruiting the Sin3 repressor complex that contains

catalytically active histone deacetylase (Jones *et al.* *Nature Genetics* 19(2):187-191 (1998)). Histone H3 and H4 deacetylation contributes to the formation of transcriptionally inactive chromatin. Thus, DMT can be used for the purpose of modulating the activity of target genes through chromatin architecture in animal cells as well as plant cells. For example, in some embodiments, DMT is used to catalytically remove 5-MeC from target gene DNA in several ways: e.g., (1) by fusing DMT to a sequence specific DNA binding protein, or (2) by fusing DMT to a subunit of the target repressor complex such as MeCP2 or Sin3. When combined with cell, tissue, or developmentally specific promoters DMT can be used to modulate specific sets of target genes.

In addition, reactive oxygen species, partially reduced species that are produced as intermediates of aerobic respiration, are powerful oxidizing agents that escape the mitochondria and attach to cellular components. Ionizing radiation and other agents that generate free radicals also produce reactive oxygen species that can attack the genome and cause lesions that are thought to have a key role in causing cancer and ageing. For example, 7,8-dihydro-8-oxoguanine (oxoG) is a very deleterious adduct generated by oxidation of the guanine base in DNA. The oxoG protein can pair with either cytosine or adenine during DNA replication. Thus, oxoG residues in DNA give rise to G/C to T/A transversion mutations. These transversions are common somatic mutations found in human cancers. HhH-GPD enzymes, such as those described herein, represent a defense against oxoG by catalysing the expulsion of the oxoG. Thus, in some embodiments, enhanced DMT activity is a method to reduce the incidence of mutations in animal cells. Also, DMT can be used to catalytically remove oxoG from a target gene by fusing DMT to a sequence specific DNA binding protein. When combined with a cell, tissue, or developmentally specific promoters DMT can be used to modulate repair of target genes.

As described above, the polypeptides of the invention can be targeted to chromosomal regions of interest by linking the polypeptides of the invention, including fragments with demethylase activity, to a DNA-binding domain that binds a target sequence. For example, it is known that an enzyme that methylates DNA (Dam methylase) can be targeted to specific sites in the genome (B.V. Steensel and S. Henikoff, *Nature Biotechnology* 18:424-428 (2000)). Specifically, the methylase was tethered to the DNA-binding domain of GAL4. When recombinant GAL4-methylase protein was expressed in transgenic Drosophila, targeted methylation occurred in a region of a few

kilobases surrounding the GAL4 DNA binding sequence. In a analogous fashion, DMT, or a portion of DMT that has biological activity (e.g., a portion containing the HhH-GpD motif amino acids such as 1167 to 1368 of SEQ ID NO:2), can be tethered (e.g., as a translational fusion or chemically linked) to proteins that interact at specific sites in the genome. As a result, specific targeted regions of the genome are hypomethylated by DMT. As discussed above, typically hypomethylation promotes transcription of genes (S. E. Jacobsen, *Current Biology* 9, 617 (1999). The invention provides compositions and methods for methylation of a desired area of the chromosome by targeting DMT to those regions. Thus, these embodiments provide additional ways to activate transcription of a desired gene in a targeted chromosomal region.

The following Examples are offered by way of illustration, not limitation.

EXAMPLE

15 **Example 1:**

This example shows the characterization of *dmt* mutant plants and the isolation of *DMT*.

Arabidopsis plants were transformed by infiltrating them with Agrobacterium containing the SKI15 T-DNA vector (generously provided by D. Weigel (Salk Institute, La Jolla, CA)). T1 seeds were harvested. The SKI15 vector has the bialaphos resistance (BAR) gene that allowed us to directly select transgenic plants in soil after spraying with the commercially available herbicide, Basta. Siliques from approximately 5,000 Basta resistant plants were opened, and those displaying approximately 50% seed abortion were identified.

25 Two lines, B13 and B33, were identified for further characterization.

Genetic analysis of the mutants revealed that the *dmt* mutants were female sterile. Male fertility, however, depended on the genetic background of the mutant alleles. For instance, in the Columbia background, transmission of the *dmt* mutation is less than 50%. However, in the Landsberg *erecta* background, transmission through the male was almost normal.

30 Molecular analysis confirmed that the two mutations were allelic. For example, both the B13 and B33 alleles carry the SKI15 T-DNA within a DMT exon, confirming that disruption of the *DMT* gene resulted in the observed B13 and B33 phenotypes.

5' - and 3' - RACE were used to delineate the 5'- and 3'-ends of the cDNA, respectively. 5'-RACE was carried out using reagents and protocols provided by 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0, GIBCO BRL, LIFE TECHNOLOGIES, Grand Island, NY and Marathon cDNA Amplification Kit, Clontech, Palo Alto, CA. Final gene specific 5'-RACE primers were SKES-4 (GGGAACAAAGTGCACCATCTCC) and SKES3.5 (CGATGATACTGTCTTCGAGC). 3'-RACE was carried out using reagents and protocols provided by Marathon cDNA Amplification Kit, Clontech, Palo Alto. Final gene-specific 3' end was obtained from cDNA library screening.

10 The nucleotide sequence of the genomic copy of *DMT* was also determined (SEQ ID NO:1). The 5'-end of the DMT RNA is located at position 3,425 of SEQ ID NO:1. The position of the 3'-end of the DMT RNA is at position 12,504 of SEQ ID NO:1. The position of the ATG translation initiation codon is at position 4,903 of SEQ ID NO:1. The position of the TAA translation termination codon is at position 15 12,321 of SEQ ID NO:1.

20 A portion of the *DMT* polynucleotide sequence, including the first exon, is encompassed by the bacterial artificial chromosome (BAC) clone T9J15TRB. For example, sequences 3820-4299, 4319-4558, 4546-5025 and 9320-9777 of SEQ ID NO:1 were previously determined using the BAC clone as a template. Moreover, a separate independently sequenced region (Bork, C. et al *Gene* 28:147-153 (1998)) also overlaps the *DMT* sequence at positions 11,087 to 12,785 of SEQ ID NO:1.

25 The predicted DMT protein has 1,729 amino acids. This sequence was compared to known protein sequences using BLAST and revealed homology to several Endonuclease III proteins. The highest homology was to the Endonuclease III protein from *Deinococcus radiodurans*, Genbank Accession No. AE002073 (see, e.g., White, O. et al. *Science* 286:1571-1577 (1999)). Other DMT motifs include two consecutive nuclear localization signals at positions 43-60 and 61-78 and a leucine zipper at positions 1330-1351.

30 **Example 2:**

This example provides further evidence that mutant phenotypes are caused by loss-of-function mutations.

A new allele, *dmt-3*, was obtained. The *dmt-3* allele was caused by insertion of the simple pD991 T-DNA vector (M. R. Sussman, et al., *Plant Physiol.*

124:1465 (2000) into the 2nd exon of the *DMT* gene. In contrast, the previous two alleles, *dmt-1* and *dmt-2*, were caused by insertion of the activation T-DNA vector, SKI015 vector. The mutant phenotypes generated by all three *dmt* alleles are the same. Because pD991 does not have activation sequences, it suggests that all three mutant
5 alleles are loss-of-function alleles. Consistent with this conclusion, seed abortion can be rescued with a transgene with 3,373 base pairs of 5'-*DMT* flanking sequences plus 1,478 base pairs of 5'-UTR ligated to a cDNA encoding the full-length DMT polypeptide (i.e., *DMTp::DMT*). Thus, *dmt/DMT* heterozygous plants that are hemizygous for the
10 *DMTp::DMT* transgene displayed 25% seed abortion. Control *dmt/DMT* plants displayed 50% seed abortion.

Example 3:

This example shows that *DMT* is necessary and sufficient for *MEA* gene expression.

As discussed above, when fertilization of *dmt/dmt* homozygous mutant flowers was prevented, fertilization-independent endosperm development was observed. This is very similar to when fertilization of mutant *mea* flowers is prevented. Thus, before fertilization, both *DMT* and *MEA*, a polycomb protein (T. Kiyosue *et al.*, *Proc. Natl. Acad. Sci. USA* 96:4186 (1999)), prevent the central cell of the female gametophyte from forming an endosperm. This is consistent with *DMT* being a positive regulator of MEDEA (MEA).

As further evidence of this relationship, *MEA* RNA accumulates in immature floral (IF) buds and open flowers (OF). However, in *dmt/dmt* mutant plants there was no detectable *MEA* RNA. Thus, *DMT* is necessary for *MEA* gene expression.

In addition, we have generated plants with a transgene, *CaMV::DMT*, designed to overexpress *DMT*. The full-length *DMT* cDNA was ligated to the constitutive cauliflower mosaic virus promoter, *CaMV* (S. G. Rogers, H. J. Klee, R. B. Horsch, R. T. Fraley, *Meth Enzymol* 153:253 (1987)). In control wild type plants, the *DMT* and *MEA* genes were not significantly expressed in the leaf. However, in
30 *35S::DMT* plants, both *DMT* and *MEA* RNA level increased significantly. This shows that *DMT* is sufficient to induce *MEA* gene expression in the leaf.

Example 4:

This example shows that DMT is a member of the HhH-GPD superfamily of DNA repair enzymes.

A BLAST search, followed by a conserved domain search, revealed that

- 5 DMT is highly related to the HhH-GPD superfamily of base excision DNA repair proteins (i.e., score of 70.1, E-value of $8e^{-13}$). This family contains a diverse range of structurally related DNA repair proteins. The superfamily is called the HhH-GPD family after its hallmark helix-hairpin-helix and Gly/Pro rich loop followed by a conserved aspartate (S. D. Bruner, *et al.*, *Nature* 403:859 (2000)). This includes endonuclease III
 10 (EC:4.9.99.18), 8-oxoguanine DNA glycosylases (i.e., yeast OGG1), the thymine DNA glycosylase of methyl-CPG binding protein MBD4 (B. Hendrich, *et al.*, *Nature* 401:301 (1999)), and DNA-3-methyladenine glycosylase II (EC:3.2.2.21). The predicted amino acid sequence of DMT contains many of the conserved amino acids of this superfamily.

The hallmark of the superfamily of base-excision DNA repair proteins is a

- 15 helix-hairpin-helix structural element followed by a Gly/Pro-rich loop and a conserved aspartic acid (i.e., HhH-GPD motif). The DMT polypeptide is 1,729 amino acids in length. Amino acids 1,271 to 1,304 correspond to the conserved HhH-GPD motif. The DMT sequence is DKAKDYLLSIRGLGLKSVECVRLLTLHNLAFFPVD. The catalytic lysine (K1286) and aspartic acid (D1304) residues are conserved in the HhH-GPD motif
 20 of DMT. Secondary structure prediction (Jpred program) indicates that DMT has two alpha-helices (amino acids 1,271 - 1,279 and 1,286 to 1,295) that correspond to the conserved alpha_K and alpha_L helices in the HhH-GPD motif of the crystallized hOGG1 DNA repair protein (Bruner *et al* *Nature* 403:859-866 (2000)).

The *Arabidopsis DMT* coding sequences were also used to identify

- 25 homologous sequences in both public and proprietary databases using both the BLAST and PSI-BLAST computer algorithms. This analysis revealed amino acid sequences from several plant species, including wheat, maize, rice, soybean and *Arabidopsis* (SEQ ID NOs: 7-29). Based on these sequences, the following consensus sequences for DMT were determined:

30 **DMT Domain A**

```
KV<1>(I, l) D(D, p) (E, v) T<3>W<1>(L, v) L(M, 1) (E, d) <0-
2>D(K, e) <1>(K, t) <1>(K, a) (W, k) (W, l) <1>(E, k) ER<2>F<1>(G, t) R<1>(D, n) (S, l) FI(A, n) RM(H, r) <1>(V, l) QG(D, n) R<1>P<1>(P, q) WKGSVVDSV(I, v) GVPLTQN(V, t) D(H, y) (L, s) SS(S, n) A(F, y) M<1>(L, v) A(A, s) <1>FP
```

DMT Domain B

W(D, n) <1> (L, f) R<5>E<3-
6 >D(S, t) <1> (D, n) (Y, w) <3>R<1>I<2>RG(M, q) (N, f) <2>L(A, s) <1>RI<2-
12 >FL<3>V<2>(H, n) G<1>IDLEWLr<2>(P, d) (P, s) (D, h) <1>(A, v) K<1>(Y, f) LL(S, e) (I, f) <1>G(5
5 L, i) GLKS(V, a) ECVRLL<1>L(H, k) <2>APPVDTNVGRI(A, c) VR(M, l) G(W, l) VPL(Q, e) PLP<2>(L, v) Q
(L, m) H(L, q) L(E, f) <1>YP<1>(L, m) (E, d) (S, n) (I, v) QK(F, y) LWPRLLCKL(D, p) Q<1>TLYELHY(Q, h
) (L, m) ITFGK<0-2>FCTK<2>PNCNAFCPM(R, k) <0-2>EC(R, k) (H, y) (F, y) (A, s) SA<1>(A, v) <0-
10 >S(A, s) (R, k) <1>(A, l) L(P, e) <1>(P, t)

DMT Domain C.

P(I, 1) (I, v) E(E, f) P<1>(S, t) P<2-5>E<0-15>(D, a) IE(D, e) <4-
23>(I, v) P<1>I<1>(I, f) (N, d) <8-17>(S, a) <1>(A, d) LV<8>(I, 1) P<2-
5>(K, r) (L, m) K<4>LRTEH<1>V(Y, f) (E, v) LPD<1>H<1>(L, i) L(E, k) <1>(D, e) D(P, i) <2>YLL(A, s
) I W(T, q) P(G, d) (B, g) <6-8>(P, s) <3>C<6-
15 10 >(M, 1) C<4>C<2>C<3>(R, k) E<5>(V, f) RGT(L, i) L<0-
22 >(L, v) FADH<1>(S, t) (S, r) <2>PI<3>(R, t) <3>(W, k) <1>L<1>(R, k) R<4>G(T, s) (S, t) <2>(S, t
) I (F, c) (R, k) (G, 1) L<1>(T, v) <2>I<2>(C, n) F(W, q) <1>G(F, y) (V, 1) C(V, 1) R<1>F(E, d) <3>(R,
g) <1>P(R, k) <1>L<2>(R, h) LH<2>(A, v) SK

- 20 The first consensus sequence listed above corresponds to amino acid positions 586 through 937 of SEQ ID NO:2. The second consensus sequence listed above corresponds to amino acid positions 1117 through 1722 of SEQ ID NO:2. The consensus sequence provides amino acid sequences by position using single letter amino acid abbreviations. Numbers in carrots (“<” or “>”) refer to amino acid positions where there is no consensus 25 and which therefore, can be any amino acid. Amino acid abbreviations in parentheses indicate alternative amino acids at the same position. Capitalized letters refer to predominant consensus amino acids and lower case letters refer to amino acids that are commonly found in DMT sequences, but are not predominant.

30

Example 5:

This example demonstrates the relationship between DNA repair and 35 demethylation.

For many years, attention was focused on the ability of DNA glycosylases to repair DNA. For example, glycosylases are involved in the repair of G/T mismatched bases by depurinating the thymidine base moiety. Recently it was shown that avian (B.

Zhu *et al.*, *Proc. Natl. Acad. Sci. USA* 97:5135 (2000) and mammalian (B. Zhu *et al.*, *Nucl. Acid Res.* 28:4157 (2000)). G/T mismatch DNA glycosylases also have 5-methylcytosine-DNA glycosylase activity. That is, these enzymes are demethylases that remove 5-methylcytosine that is later replaced by cytosine. Without intending to limit the scope of the invention, it is believed that as a member of this superfamily, *DMT* is a demethylase (i.e., 5-methylcytosine glycosylase).

The methylation (i.e., amount of 5-methylcytosine) state of a gene can have a profound effect on its expression. In general, hypomethylation is associated with elevated gene expression, whereas hypermethylation is associated with decreased gene expression (S. E. Jacobsen, *Current Biology* 9:617 (1999)). Thus, it is likely that *DMT* activates *MEA* gene expression by reducing its level of methylation.

Mutations in the *DDM1* gene in *Arabidopsis* reduce by 70% the overall genome cytosine methylation (E. J. Finnegan, *et al.*, *Proc. Natl. Acad. Sci. USA* 93:8449 (1996); M. J. Ronemus, *et al.*, *Science* 273:654 (1996)). Such plants develop a number of phenotypic abnormalities including floral phenotypes (T. Kakutani, *et al.*, *Proc. Natl. Acad. Sci. USA* 93:12406 (1996)). Similarly, phenotypic abnormalities have been observed developing in *dmt/dmt* homozygous plants that affect petal number, floral organ fusion, and floral organ identity. Moreover, independent *CaMV::DMT* transgenic lines that overexpress *DMT* frequently are late-flowering. This is particularly interesting because late flowering of *ddm1* plants was shown to be due to hypomethylation of the *FWA* gene (W. J. J. Soppe *et al.*, *Mol Cell* 6:791 (2000)). Thus, without intending to limit the scope of the invention, it is believed that both *ddm1* loss-of-function mutations and overexpression of *DMT* (i.e., *CaMV::DMT*) may result in genome hypomethylation.

25 **Example 6:**

This example demonstrates targeting gene expression to the female gametophyte using a *DMT* promoter sequence.

DMT RNA accumulates in many plant organs such as immature flowers, mature flowers, open flowers, stems and to a lesser extent, leaves. To understand the spatial and temporal regulation of *DMT* RNA accumulation, the expression of the *DMT* promoter fused to reporter genes was analyzed. We fused 2,282 base pairs of 5'-*DMT* sequences, the full-length 5'-UTR (1,478 base pairs), 444 base pairs of *DMT* coding sequences that contain a nuclear localization signal to two reporter genes, the green

fluorescent protein (*GFP*; (Y. Niwa, *et al.*, *Plant J.* 18:455 (1999))) and β -glucuronidase (*GUS*; (R. A. Jefferson, T. A. Kavanagh, M. V. Bevan, *EMBO J.* 6:3901 (1987))). Reporter gene expression was observed in the developing female gametophyte, in the polar nuclei before they fuse, in the egg and synergids, and in the central cell. Expression 5 was not detected after fertilization. Thus, this promoter is useful for targeting gene expression to the female gametophyte.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof 10 will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

SEQ ID NO:1 DMT genomic sequence

DMT genomic sequence (12,785 bp)

AAGCTTAAAGCTACCAACATCGAATTAGTAAAAGACCCATGATTGAAATTGGAATTGTCGG
 5 CAAATCGAGAAGATAT
 AGAGCCGACACGGGAACAGTGAAAACCACAAAGCGCGTAAGAACATGAAACAGTGGGAGAAGG
 AAGAGAGAACATTTACCGAT
 CATTGAGGGAAAAGATGGGAATCAGAGAAAAATCTGGAAAAAAAGAAATTAAAGAGAAAGA
 GAGAGAAGAAAGTGAGGAG
 10 GAAGATGCAGTGAAGACTGTATGCCACATCCCACATGGTGTGATGAGAGAGAGAGAGAGA
 GAGGTTAACGAGCAAAT
 TGTGGAGAGATAAAAGAGAGAGAGAGACTGAGCGAGTCAGTCGTCGTGTTAAAAGAA
 AGAATCCTATATTGCTT
 15 TTTCCTTACTACTTATTTCAAGACTATTGCTTATTGCTCAAACCTTTTGATTTGACTT
 TTGAGATCCTAAAGT
 GTTGTACAATITACCTGCTTTCTCCAAGAAAAATCAGAACAGACCACAGCAAATTATGTA
 TTTCTATTAAAAAAAG
 AAAGAAAGAACATTATTAAGATAATTAAAGCTAACAGTTGAAACGTGAAAGCAGA
 ATTTCTAAAAAAATAGT
 20 AAACTGCTACAAACTTATTATGTGTATATAACATATCTATAAAGAAACTCAAATATGATA
 AACATTTAACAAAT
 TTCTATGAAATTATAATAAAAAAAAGTCACTTTGACACTTAAAGGTTGACAATAACCGTCTCT
 CCAAAAAAAATCAA
 ACATTTATAATTCTAAAACATGGTGTATTTGCTGAAATCAAAGAAAAGAAGGATTTC
 25 TATATCATAAGTTCAT
 TATTGTATCCTAAACTTCAAAATTTCATGTAATTGAAAGGAAAAATTAAAGATATAATGTTGTT
 TTGTTTCTTATGTTA
 CATTTCTGGAATATATATTCTAAACAAAAATGTATTAAATATGATGAGAGATTACCATCC
 AAAAGGTGCAACTTAT
 30 ATAAAACAAGTTAATAACTAAACAATACATGTGATCACATCAATGACAGTTGATCTTAA
 ATAGAAATGATTGAGCA
 AACCTCAAAATGTCCTTAGGATCACAAATCTTCCTTAGCTTAAAGCCGGGAGTT
 AACTCTCTCCCTTG
 TAGACTTTGTTCTAAATCTTTCTTCAAAAAATCAATAATTAGTTAATGGGCATAATATT
 35 TGGTTTAAATTAGT
 CCATAGATTTCAGGACCATCTCTAAATCACGACAAATATCTAAATTGTAACACATTAAAC
 TTAAAAGTATTGCTT
 CACAATCCTTAAATATATATATATATATATATATATATATATATATGAAAGTT
 ATAGAAACGATAACTC

CTTACTCAACAATTAGCCAAAAAAACATCCATAATGCATTTAACTAGGAATTITAACAAAC
 TCAAATAGGGTAGT
 TAAAAAAAAACAAATAGTAGATGTACATACGTACCTTTAAAATATACTCATATCGAAAGT
 TTAAATTGCGAAAT
 5 TAAATACATTATCTATAATTAAAATACATTTAATAATGCATAATTCTGTAATATCTATCTT
 AATTCCATATAGAAC
 CAAACAAAATAACATATCAAATAGTTAACCTAACAAAAACGTTAGGGAAAGTTGACCT
 AACTAGCTGATTGACG
 TTGAACTTGTCAATGCGAAAAGCGATATTCCAATATATACTACATGTAGTATTATTTATGG
 10 AGTTCTAAAAGGT
 TTGAGTGGATTGTTACTTGTGGAGGATGCTATTTCCTCTGCCATAATATTACGAGTA
 TGGGATAACTACATA
 CTCATGATTATGAAACGCTACTTTATTGAAAACCTCTTAACACCAAATATGTCACTAGA
 TTCCAAAACGTTAGACC
 15 AATTGTATCTAATCTCAAATTCTCAATCAAAGTATTAAATTACCGATGGTAAGAAAAGTTAAC
 GATATAATTATCAAAA
 GAAAGAATAAGTCAACAGATTCTTAATCTTTATTGGTATATGAACATTGTACAAAATC
 TCAAAAGATATGTAAAC
 20 TGTTTAAATATAAATTCACTGAGATTAATTCTCAGACTCGTGTAGCTATAATAATGTCAAG
 AGTTCTCTGTTCA
 AGGAAAACCTTAAAGATATGTATATTCTGTAATTATGATGATATAATTGCTATTCTGT
 CACAAACATTACTTA
 AAAATCGTATTCTCATTACTACAATGTTGACTAAGAACAAAATACATTGATTATTGATATAT
 CGTCAACTGAATTTC
 25 TTCCGAGGGATAATTCTCAAACATAGCAAGAACATCTCATAATAATTGTTCGTACTACCTTA
 GAGCAAATTITTTAA
 GATTGTAACGTGACTTATGGCTCTTGCTGGGGTCAATGCGAATAATCTAAATGTATG
 GGAGTCAAATAAAATAC
 CAAGAAAATAAGGAGCAGCACCCAATAAACTATATGGGACCAGAAATCCTTCATTGGTT
 30 AAAATAGGATTATCCG
 AAAGATGAAGGACTAAATTGAAACTGATTGGGGTAGGAAGAGATCCGTACAATCTTAAT
 GGCTTCCACGCGGAAACT
 TGTCGTTATACAATTCTTACGAGCTGGGTTATATTCAAATGGTCAAATAATAT
 TAGTTAACACACTAA
 35 CGGAGTAATTAAATTGGTACTACAATTCTCAGTTGGTGCATTAGAACGAAACATAGTCG
 TAAAATACGAGTCGGT
 GTTATACCTTATTACGTTAAAAAAATACGAGAATTGTTGTCAATTCTAAATTATTCAT
 GAATATATGGAAATTAA
 TTAGATACTCTAGCGAAAATAGTGATTATGAGCGTTACAAAATACGATTTCATTGAA
 40 CTTCCATTATGTAATTCTCCTTATGTAATTCT

GGTCAAATGTTGGCATGAAGAAGCAAGTTGCAACATTAATTCATTAAAAATCGTGTGA
 CATACTTAAATCTAA
 ATATAGGAAGAACCAAAACATTAATTTAGTAAGATCTAATGAACATTATAAGTTATAA
 CTTATAACCAACAAAG
 5 TTGGGTTAGCGTTGTTCTTATCTGAAAACCTGCAAACCTAACCCATTAAAGGACTAATG
 ACAATTAAACAACAAAG
 TACACTTAAGCAACACGCCTCGTGAATATAATTGGCCTCAGGCCATATTGCTAACGCC
 AACTGATATTCACTTT
 ATTCCCTCTCATCTCACACACTCTCTCTATCTCTATCTCTAAGGCATAGCTGACTCAGTG
 10 TTCTCCGGCATGAC
 TCGCCTGAGAATCAGAAAGCTTAGATCGGTGAGCTTTAGCTCCATTCTGTTATTACATA
 TTATTTCTTTTTTC
 TCTCTCCCTTTTATCTGGAATTGTTGCTAAATTTCAGCTGTTACATTCCGATCACG
 AGAAGAACACTGGG
 15 TTTTATGTTAACATCAACATGTTCTGTTCTGATCATAAATCTCAGCTATTACACCTGATT
 TTGATTCTGCGTAAT
 AAAAACCTCTGATTGCTTTATCTTCACTTCCCCATAAACATTGCTTACTTTATTGCTCTTC
 TTTTACCGTTCCAG
 CTAAAAAAATTCTCGTATTCAATGTGTTCTGTTGATGAGAAAAATATCTGACAAAAA
 20 AACATTTATTGCATT
 TTATGGTGAGATTCTTAGTTAACATGTCGCTCTCAACCAAGTCAGATTAAAAGGAGTGTTC
 GTCCATGTGCTTTGT
 TTGGTGTGGAGAGAGTTCCGGAGAGTTAGGTGAGTGTATTGGGGTGAGGTAGTGATA
 AGGGTTGAAGGGGGAGT
 25 GATTCACTAACGTGTTATGAATTGAGGGCTGATCCGGGGATAGATAATTGAGTTCCCTT
 GGAGAACAAACTCAA
 CAAGAGTTCAATGGGTTCTGGATTCCATTACACCCAAAAACCTAGATCAAGTCTGATGGTA
 GATGAGAGTGATAAA
 CCAGGATCTAAATGGGTTCCAGGTGGTGAATTGAGACAGGGATTCTGCAACACTGGTGT
 30 GGATCATAATGGGGTTT
 TTGATCATGGTGTCTCAGGGCTTACCAACTTAAGTATGATGATCAATAGCTAGCGGGAT
 CACATGCAAGCTTGG
 AGTAATACTGAGAGAGATCTTTGGCAGGAGTGAGGTGACTTCTCCTTAGCACCAGTTATC
 AGAAACACCAACCGGTAA
 35 TGTAGAGCCGGTCAATGGAATTCTCAGATGTGGGTATGGTAATGGCCTTCACCCA
 GAGTGGCAGTCTCAAG
 CTGGCTATAATGAGTTGAAATTGGATGACTTGTGAATCCTGATCAGATGCCCTCTCCTICAC
 AAGCTTGTGAGTGGT
 GGGGATAGCTTATTCAAGGTTCGTCAATGTGAGTGATCAAATCTATTTCAGTTTTTTTCC
 40 CTTTCITCCGTTCTT

GCAGTACTTAGAGTAGAACATGAATTAGAATATCTTAAGAAAGTCATGGTTTGAAACAGATGG
 ACCTCCACCGGTGAAACA
 AGCCTTCAATTGAAATTCACCAATTAGAAGAGAACAGCTGGTCAGTCGTGAAAGTT
 CGTTCAATATGTACCG
 5 TCAACGCCAGTCTGTCAGAACACAGGTGAAAAGACTGGATTCCCTGAAACAGATAGTTACA
 ACTGGACATGAAATCCC
 AGAGCCGAAATCTGCACAAAAGTATGCAGAGCATTATGGACTCGTCTGTTAATGCGACGGA
 AGCTACTGAACAAATG
 ATGGCAGCAGACAAGATGTTCTGGAGITCTGACCTTAACAAAACCTCCTCAGCAGAAACCC
 10 AAAGGAAAAGGAAGTTC
 ATGCCCAAGGTGGCTGGAGGCAAACCTAAAGAAAGCCACGCAAACCTGCAGAACITCC
 CAAAGTGGCTGGAGG
 CAAACCTAAAAGGAAGCCACGCAAAGCTGCAACTCAGGAAAAAGTGAATCTAAAGAAACCG
 GGAGTGCCAAAAGAAAA
 15 ATTTGAAAGAATCAGCAACTAAAAGCCAGCCAATGTTGGAGATATGAGCAACAAAGCCCT
 GAAGTCACACTCAAAGT
 TGCAGAAAAGCTTGAATTTCGACTTGGAGAATCCTGGAGATGCGAGGCAAGGTGACTCTGAG
 TCTGAAATTGTCAGAA
 20 CAGTAGTGGCGCAAACTCGTTCTGAGATCAGAGATGCCATTGGGAACATAATGGTAGTT
 CCTGGATTCACTGTCA
 AAATAGACAAGACCAATGGATTGGGGCTATGAACCAGCCACTTGAAGTGTCAATGGAAAC
 CAGCCAGATAAACTATCT
 ACAGGAGCGAAACTGGCCAGAGACCAACACCTGATTATTGACTAGAAACCAGCAATGCCA
 GTTCCCAGTGGCAACCCA
 25 GAACACCCAGTCCCAATGGAAAACCAACAAGCTGGCTTCAGATGAAAACCAACTTATTGG
 CTTCATTTGTAACC
 AGCACCTCGCATGACCATAAAGAAACCAGCAGCCTGCTGGCATGGTAATCAACAAACCTA
 TGATCTGATAGGAAC
 CCACGGCCTGCATTAGTAAGTGGAAACAGCAACTAGGAGGTCCCCAAGGAAACAAGCGGCC
 30 TATATTTGAAATC
 GACTTGTACCTGCTGGAAATCAGCTATGGATCACCTACAGACATGCATCAACTTGTATG
 TCAACCGGAGGGCAAC
 AACATGGACTACTGATAAAAACAGCAACCTGGATCATTAAAGAGGCCAGCAGCCTG
 GTACCTTGTGACCAAG
 35 CAACCTGCAACTCCAAAAGGTTTACTCACTTGAATCAGATGGTAGCTACAGCATGTCATCG
 CCTGGGCTCGACCTCA
 TTCTCAGTCACAAGTCTCACAACATATCTACATGTGGAATCTGTTCCAGGATTTGAATGGG
 ACTACAGGTACATGCC
 AGAGAACAGGGCTCTGCATACGATTCTTACAGCAAGATATCCATCAAGGAAATAAGTACA
 40 TACTTCTCATGAGATA

TCCAATGGTAATGGGTGCAAGAAAGCGTTACCTCAAAACTCTCTGCCAACCTCAATTATG
 GCTAAACTTGAGGAAC
 CAGGGGCTGAAGAGACAGTATCATCGTCAATGGGACAGACGGAAAAGCATGATCTAAACT
 TAGCTCAACAGATIGCTC
 5 AATCACAAAGATGTGGAGAGACATAACAGCAGCACGTGTGGAATTTAGATGCTGCAAAG
 AAAACGAAAATCCAGAAA
 GTAGTCCAAGAAAATTGATGGCATGCCACCTGAGGTATAGAAATCGAGGATGATCCAAC
 GATGGGGCAAGAAAAGG
 TAAAAAATCTGCCAGCATCGTAAAGGTGCATCTAAAGGAACTCGTCTCCAGTAAAGAAC
 10 AGCAGAAAAGGAGAAAAT
 GTATTGTCACAAACGCCGTGCAAAAAAGGGTCGAGCAGGTAGAAAAAAATCAGTACCTCCG
 CCTGCTCATGCCCTAGAG
 ATCCAGCTTGGCACCTACTCCTCCAAAGACACCTTATCAAGAAGCAAGCCTAAAGGAAA
 GGGAGAAAGTCCATACA
 15 AGATTCAAGAAAAGCAAGAGGTAACTAATGTATTCTACAATCTCTGTGATATAATTTGAGAT
 TTAGTAACTGATGTG
 CCAAACCGACTCCTTATACTGTTGGTGCCTGTATAGGTCCATCAGGAGAACCTCTGTGTCAG
 GATTCTATTGCGGAAA
 TAATTTCACAGGATGCAAAATCTGTATCTAGGAGACAAAGAAAGAGAACAGAGCAAAATGCA
 20 ATGGTCTTGACAAAGGA
 GATGGTGCACTTGCCCTATGAGAGCAAGAAGCGAAAACCAAGACCCAAAGTTGACATTGA
 CGATGAAACAACCGCAT
 ATGGAACCTACTGATGGGAAAGGAGATGAAAAGAAGGGATGAAGAGAACGGATAAAAG
 AAAGAGAACAGTGGTGGAAAG
 25 AAGAAAGAACAGACTTCCCGAGGAAGGGCTGATTCTCATCGCTCGCATGCACCTGGTACAAG
 GTGAAGATCCACTTCTC
 TCTCAACTCCATTTCACACAAATTAGTAGAATACTCAAAATGATGTTTGTTGCAA
 AATTAAATTCACT
 AGTTAACCATGTCACAAATAATTCTATAATGCATTTGTGAAGAACAGGTGTGATTTATGGT
 30 ACAGCTGAATGGTTAT
 GTGCCATTATTTCTTACTGCTATAGATGACCAATTGAACTTAAACGTTACAGGAGATAGA
 CGTTTTGCCATGGA
 AGGGATCGGTGGTGTGATCGGTATTGGAGTTCCCTACACAGAACAGTCTCGGATCACCTTC
 AAGGTATATGAGITGC
 35 CTTAATAAAATTGAGTICCAAAACATAGAAATTAAACCCATGGTGGTTTACAATGCAGCTGC
 GTTCATGTCCTAGCTG
 CTCGATTCCCTCCAAAATTAAAGCAGCAGCCGAGAACAGATGAAAGGAATGTTAGAACGGTAGTT
 GTGAAGATCCAGAAGGA
 TGCATTCTGAACCTAAATGAAATTCCCTCGTGGCAGGAAAAGGTTCAACATCCATCTGACATG
 40 GAAGTTCTGGGGTTGA

TAGTGGATCAAAGAGCAGCTAAGGGACTGTTAACTCTGGAAATTGAAAGATTAAATTCTT
 AGAGAAGAGTATTCAAA
 ATTTAGAAGAGGAAGTATTATCATACAAGATTCTTGTATCGGCATATTCAAGTCGTGTGG
 GAGAGTTGGATCCTGT
 5 TCATGTTCCAATCAGACGCAGAGTTCTACAACCAGGTGTAAACAAAAACTGTCAGTGGAA
 ACATCACAACTCAGTGCA
 AACTGGGAGCCCAAATTGTCGATGAAATTGTCCTCAAGGGAAATGAGAGACCGCATTATA
 TGAAGGATCTGGTGTG
 TTCAGAAACAAGAAAATACAATGTCGCTCAGAAGAACCTGATCTGAAAAAAACATGAAT
 10 TGAAAGACTCTGTCGT
 TTGGTCAGCCAAGAAATGATACTAATTGCAAACAACCTCCAGCAGCTATGAGCAGTGT
 GCGACTCGACAGCCCAA
 TGTACTAGACATAGAGGATTGGAATGCAGGGTAAGGCCCTGGTTATTCTGGATGTCCAT
 CTCACCAAGAGTTGACA
 15 GAGTAAAGAACAAAAATGTACACGCAGGTTTCAGACAAGGTGGAAGTGTCCAAGAGAA
 TTCACAGGTCACTACATA
 CCATCAACGCCTCATGAATTACCAAGGAATGGATTGTCGGTTCTCAAGGCCGTCCAAGAA
 CACCAAGGAGCATACCCA
 ACATAATCAACAAGATGAGATGAATAAACGATCCCATTACAAAAAACATTGGATCTGCT
 20 CAACTCCCTGAAGAAT
 GCCTTACAAGACAGTCAGTACCAACAGAACATCACGGATGGCTGTACCGAGAGATAGA
 ACTGCTGAAGACGTGGTT
 GATCGCCTCAGTAACAATTCAAGCTTACAGAACATATTGGTCGAATCAAATTCCAGCAATAAA
 GAGCAGACGGCAGTTGA
 25 ATACAAGGAGACAAATGCCACTATTACAGAGAGATGAAAGGGACGCTGCTGATGGGAAA
 AGCTTACAAGGCCAGTGGG
 ATAGTCTCAGAAAAGATGGAGGGGAATGAAGGGAGACAGGAACGAAACAAAAACAATAT
 GGATTCATAGACTATGAA
 GCAATAAGACGTGCTAGTATCAGCGAGATTCTGAGGCTATCAAGGAAAGAGGGATGAATAA
 30 CATGTTGCCGTACGAAT
 TAAGGTTAAATCTACTAATTCAAGTGTGAGACCCATCAAACTGTCAGAAGGCTGAAACATCA
 GTAAATTATGTAACCAT
 ATTACACATTGCAGGATTCTAGAACGGATAGTTAAAGATCATGGTGTATCGACCTGAA
 ATGGTTGAGAGAATCTC
 35 CTCTGATAAACGCCAAGTGGTAAATCACATTAGTGAACACTGCAACACTAGCACGATCGATT
 TACTCAACAATTACGTC
 AAACTGAGTATTAACAAGATTGCTCATGAACATTCACAGGGACTATCTCTGAGCATAAGAGG
 TCTGGGTTGAAAAGTG
 TTGAATGCGTGCAGCTTAACACTCCACAATCTGCTTCCCTGTGAGTCAGACTATTCCATT
 40 ATCTACTAAAAACTTA

GAATAACTCCGGCTAACTAAGCTGGAACTTGTATTGATGATATGAAGGTTGACA CGAATGTTG
 GAAGGATAGCAGTAGG
 ATGGGATGGGTGCCCTACAACCCCTACCTGAATCACTTCAGTTACACCTCCTGGAGCTGTAA
 GTTCTTTTGTGTC
 5 ATCTAAACACGAAATTCTATGCAAGTCATAACCAGCTGTGTTTACAGATAACCGAGTGCT
 CGAGTCATCCAAAAA
 TTCTTGGCCAAGACTTGCACACTCGATCACGAACACTGTATGCTCATAAACTCTAACAAA
 TCATCTGCTGAAAAA
 CCAATATTCTTGGTAGAATTCTATIGCTATTACTCATTACTAACAGCGAAATTAAACGTT
 10 TCTTTTCTTACTCGA
 GTATGAATTACACTACCAACTGATTACGTTGGAAAGGTATTATGCTCTAACGCTTGAATTAA
 TCATATGGTAATTCA
 AGCATTGTAGGCACCTGATCAATTATGTGCTAAATCATGTGAATTATGTCAGGTATTGCA
 CAAAGAGTAGACCAAA
 15 TTGTAATGCATGCCAATGAGAGGAGGTGCAGACACTTGCAGTGCTTATGCTAGGTAAGC
 AAGCTTCATGACTTA
 TATGCAATAATTAAAGATAAAATTAGGATTATGGTAAGTTACAAAAAATTAGGCTCAGTT
 CATGGTAGCTGGAA
 AATAGTATTACAAGAACACATAAAGATCAAAGACAGAACATGGATCCATATGCACTATCAT
 20 TTAGCTCTGTAAATCC
 ATACATGAACACTATATGCCAAGTAGGGATTCAAATATGAGATTGACTGATGCCATT
 GTAACAGTGCAGACTT
 GCTTACCGGCACCAAGAGGGAGGGAGCTAACAGTGCAACTATTCCGTCCTCCGAGTCC
 TATCCTCTGTAGGCCAT
 25 CCCGATGATAGAACACTCTTCCGTTGGAGAAATCCCTAGCAAGTGGAGCACCATCGAACATAG
 AGAAAATCTGTGAAACCA
 TAATTGAAGAGCCGCCCTCGCCCGGGCAAGAGTGCACTGAAATAACCGAGAGTGATATTGAA
 GATGCTTACATAATGAG
 GACCCCTGACGAGAACCAACAATAAAACTCAACATTGAAACAGTTGAAATGACTCTACGGAA
 30 CACATGAAAGAACAT
 GGAGCTCCAAGAAGGTGACATGTCCAAGGTTGGTTGCTTGCATCCAACAACTACTCTATT
 CCAACTCCCAAACCTAA
 AGAACATTAGCGCTCAGGACAGAGCACCAAGTGTAAAGCTAATATCCTCTATATTATC
 TTCCATATAAATTG
 35 GGGAAAAAAATCGCTCTCCATCTGGTTTAGAACATGCGGGTCAGCCAGGGTTATGGCATT
 ATATATTCTACCGATCG
 GCCCGAGCTGGCTCTGGTTGACTCGTATGCCACCTGCATTGAAACAAACAGTAGGAGACAAG
 CAAGCAAAACGTTTAA
 GATAAGGCTATGGTAAATGACAAGGTAACGTGATAATGTGTCGTCTATTGCAAGGTACGAG
 40 CTCCCAGATTACATCG

TCTCCTTGATGGTGAAGTCATTAACTCTCTATACTCGAGTTTCACTTGCAGCAACA
 CTGTTAAAAGTCCT
 CATTGATAAAATAACAGATGGATAAAAGAGAACCGAGATGATCCAAGTCCTATCTCTAGCT
 ATATGGACACCAGGTGA
 5 GAATAAAAATGCAATGTTCATGTCAGTACAGTATCAAAGAAAGTACAGCTAGAGCTAA
 AAAGCATTGAAATAGA
 GTCGGTTAAATATGAAAGTTGAATCTGAAATGAAAGCCGGAACGTAGCATTGGTGGATGTT
 ATATGTAATTAGTTT
 TGAGATTGGCTAAATGTAATTGACTGTCAGCAGGTGAAACAGCGAACCGGCCT
 10 GACAGAAGTGTGGAGG
 GAAAGCGCTGGCAAAATGTGCTTGCAGGACTTGTCTGAGTGTAAACAGTCTGAGGGAAAGC
 AAACTCACAGACAGTTC
 GAGGAACCTCTGGTGAAGATTCTGATCTTGTGCTCATGAAAAGGAGAAGTGAGA
 ATACAAGTTGCTAATA
 15 TCATTITTCGTCATTCAAGATACTTGTGGACTGCCATGAGAGGAAGTTCCGCTCAACG
 GGACATATTCAGT
 CAACGAGGTTAGATGAAATAAAACTCAAACAGACAGACGAAACATTATTCTGTTAGTGTG
 GTTCTTATCCTCCCTG
 CCATTTTTATCTGCAGTTTGCAGACCACGAGTCCAGTCTCAAACCCATCGATGTTCTA
 20 GAGATTGGATATGGGA
 TCTCCCAGAAGGACTGTTACTTCGGAACATCAGTAACATCAATTAGAGGTTAAACAT
 TCGTAATAGAGTTAGT
 AATCAAATGTCAAAACAAGAAAGCTCACCGTCAAACACAAGAAAGCTCACCTCTC
 TTGCCAAAAAAAGATCT
 25 TAGAATGTTTGCTGAATTGTCAGGTCTTCAACGGAGCAGATACTGTTGCTTTGGAAA
 GGTAAACGTTAACTT
 CGACCCAGAGAAATCGGAAAATCTATTGCTTGTCTGATCAATACGTTAAACATACACA
 CACACTTACACTTAGG
 ACCAATACTGTTCTGATCTGATAGAAACTGGTAAACATCTAACAAATTGATTGCAGGATT
 30 CGTATGTCCTGGAT
 TCGAACAGAACAGAACAGCACCCTCCATTAAATGGCAAGGTTGCATTTCCTGCGAGCAAAT
 TGAAGAACACAAAACC
 TAAAGATGACTGGAAGAAAGCAAACGCATTGCTCTGCTCCTCTATTAAAGCCAGGAA
 AAGTCCCATTAGACAT
 35 AATAACAGGAATCCAATAGGCTATTCTCTTCTTCTTATTICATTCAAGAGCAGAACGG
 ACACAAAAAAAGTTTT
 TGGGTTATTATTTCTCTCTAAACAAATTGTCAGCGTTGGGTCTTTCTGGCTGTCACTAGC
 GTGGCAAAATCCAATG
 TCCGCGCACACTTAGGCGCATTGCAATAAAATTCTCCGGCCACGGAGTGTACGATTTTCC
 40 AACGGCGGCTAATGCG

ATATTTCCGGTAACACATATTCTTATCTATGTTGGTTTGTCAGCGTGGGCCTACTAG
 ACAATGATCATCAATA
 AAACTAACACAAAGTTGAATGCTACAAAGTAGAAAGTGAAGAAAAATAATAGACATTGC
 CGA

5

SEQ ID NO:2 DMT amino acid sequence

MQSIMDSSAVNATEATEQNDGSRQDVLEFDLNKTPQQPKPSKRKRKFMPKVVEGPKPKPRKPA
 ELPKVVVEGPKPKRPR

KAATQEVKVSKETGSACKNLKESATKKPANVGDMNSNSPEVTLKSCRKALNFDLENPGDARQG
 10 DSEIEIVQNNSGSNF
 SEIRDAIGGTNGSFLDSVSQIDKTNGLGMAMNQPLEVSMGNQPDKLSTGAKLARDQQPDLLTRNQQ
 CQFPVATQNTQFPME
 NQQAWLQMKNQLIGFPGNQQPRMTIRNQQPCLAMGNQQPMYLIGTPRPALVSGNQLGGPQGN
 KRPFLNHQTCLPAGN

15 QLYGSPTDMHQLVMSTGGQQHGGLIKNQQPGSLIRGQQPCVPLIDQQPATPKGFTHLNQMVATSM
 SSPGLRPHSQSVPT
 TYLVHEVSRSRILNGTTGTCQRSRAPAYDSLQQDIHQGNKYILSHEISNGNGCKKALPQNSSLPTPI
 AKLEEARGSKRQY

HRAMQTEKHDLNLAAQQIAQSVDVERHNSSTCVEYLDAAKTKIQKVVQENLHGMPPEVIEEDD
 20 PTDGARKGKNTASIS
 KGASKGNNSPVKKTAKEKEKCIVPKTPAKKGRAGRKKSVPPAHASEIQLWQPTPPKPLSRSPKG
 KGRKSIQDSGKARG
 PSGELLCQDSIAEIYRMQNLYLGDKEREQEQQNAMEVLYKGDGALVPYESKKRKPRAKVDIDDETTR
 IWNLLMGKDEKEG

25 DEEKDKKKEKWEEERRVFRGRADSFIAARMHLVQGDRRFSPWKGSVVDSVIGVFLTQNVSDHLS
 SSAFMSLAAFPKKLS

SSREDERNVRSSVVEDPEGCILNLNEIPSWQEKVQHPSDMEVSGVDSGSKEQLRDCNSGIERFNFL
 EKSIQNLEEVLS

SQDSFDPAIFQSCGRGVGSCCSKSDAEFPTRCETKTVSGTSQSVQTGPNSLDEICLQGNERPHLYE
 30 GSGDVQKQETTN

VAQKQPDELKTMNWKDSVCFGQPRNDTNWQTPSSSYEQCATRQPHVLDIEDFGMQGEGLGYS
 WMSISPRVDRVKNKNVIP

RRFFRQGGSVPREFTGQIIPSTPHELPGMGLSGSSSAVQEHQDDTQHNQQDEMNKASHLQKTFDL
 LNSSEECLTRQSST

35 KQNITDGCLPRDRTAEDVVDPASNSSLQNILVESNSSNKEQTAVEYKETNATILREMKGTLADKG
 KPTSQWDLSRKDVE

GNEGRQERNKNNMDSIDYEAIRRASISEISEAIKERGMNNMLAVRIKDFLERIVKDHHGIDLEWLRE
 SPPDKADYLLSI

RGLGLKSVECVRLTLHNLAFPVDTNVGRIA VRMGWVPLQPLPESLQLHLLELYPVLESIQKFLWP

40 RLCKLKDQRQTLYELH

YQLITFGKVFTKSRPNACPMRGECHFASA YASARIALPAPEERSLTSATIPVPPESFPVVAIPM
 IELPLPLEKSLA
 SGAPSNRENCEPIIEEPASPGQECTEITESDIEDAYNEDPDEIPTIKLNIEQFGMTLREHMERNMELQ
 EGDMSKALVAL
 5 HPTTTSIPTPKLNISRLRTEHQVYELPDSHRLLDGMDKREPDDPSYLLAIWTPGETANSAQPPEQ
 KCGKGASKMCFD
 ETCSECNSLREANSQTVRGTLIIPCRTAMRGSFPLNGTYFQVNELFADHESSLKPIDVPRDWIWDLP
 RRTVYFGTSVTI
 FRGLSTEQIQFCFWKGFCVVRGFEEQKTRAPRPLMARLHPASKLKNNKT

10

SEQ ID NO:3 DMT 5' flanking sequence

AAGCTTAAAGCTACCAACATCGAATTAGTAAAAGACCCATGATTGAAATTGGAATTGTCGG
 CAAATCGAGAACGAT
 AGAGCCGACACGGGAACAGTGAAAACCACAAAGCGCGTAAGAATGAAACAGTGGGAGAAGG
 15 AAGAGAGAACATTTACCGAT
 CATTGAGGGAAAAGATGGGAATCAGAGAAAAATCTGGAAAAAAAGAAATTAAAGAGAAAGA
 GAGAGAAGAACGTGAGGAG
 GAAGATGCAGTGAAAGACTGCTATGCCACATCCCACATGGTGTGATGAGAGAGAGAGAGA
 GAGGTTAAAGCAGCAAAT
 20 TGTGGAGAGATAAAAGAGAGAGAGAGACTGAGCGAGTCAGTCGTCGTGTTAAAAGAA
 AGAATCCTATATTGCCT
 TTTTCTTACTACTTATTTCAGACTATTGCTTATTGCTCAAACCTTTTGTATTGTCACTT
 TTCGATCCTAAAGT
 GTTTGACAATTACCTGCCCTTTCTCCAAGAAAAATCAGAACAGACCAAGCAAATTATGTA
 25 TTTTCTATTAAAAAAAG
 AAAGAAAAGAATTCAATTACTTATAGAATTAAAAGCTAAGCAGTTGAAAACGTGAAAGCAGA
 ATTTCTAAAAAAATAGT
 AAACTGCTACAAACTTATTATGTGTATATAACATATCTATAAGAAACTCAAATATATGATA
 AACATTTAACAAAT
 30 TTCTATGAAATTATAATAAAAAAAAGTCACTTTGACACTAAAGGTTGACAATAACCGTCCT
 CCAAAAAAAATCAA
 ACATTTATAATTCTAAAACATGGTGTATTTGCTGAAATCAAAAAGAAAAGAAGGATTTC
 TATATCATAAAGTTTCAT
 TATTGTTCTAACTTICAAATTICATGTAATTGAAAGGAAAAAAATTAGATATAATGTTGTT
 35 TTGTTCTTATGTTA
 CATTTCTATGGAATATATATTCTATAACAAAAAAATGTATTTAATATGATGAGAGATTACCATCC
 AAAAGGTCGAACCTTAT
 ATAAAACAAGTTAATAACTAAACAATACATGTGATCACAAATCAATGACAGTTGATCTTAAA
 ATAGAAATGATTGAGCA

AACCTCAAAAATGCTTCTAGGATCACAAATCTTCCTTAGCTTATTAAAGCCGGGAGTT
 AACTCTCTCCCTTG
 TAGACTTTGTTTCAAATCTTCTTCAAAAATCAATAATTAGTTAATGGCATAATATT
 TGGTTTAATTAAAGT
 5 CCATA GATTTT TAGGACCATCTCTAACGACAAATATCTAAATTGTAACACATTAAAC
 TTAAAAGTATTGCAIT
 CACAATCCTTAAATATATATATATATATATATATATATATATATATATATGAAAGTTAT
 ATAGAAACGATAACTC
 CTTACTCAACAAATTAGCCAAAAAAACATCCATAATGCATTAACTAGGAATTAAACAAAC
 10 TCAAATAGTTGGTAGT
 TAAAAAAAAACAAATAGTAGATGTACATACGTACCTTAAATATACTCATATCGAAAGT
 TTAAATTTGCGAAAT
 TAAATACATTATCTATCAAATAAACATTTAATAATGCATAATTCTGTAATATCTATCTT
 AATTCCATATAGAAC
 15 CAAAACAAAATAAACATATCAAATAGTTAACCTAACAAAACGTTAGGGAAAAGTTGACCT
 AACTAGCTGATTGACG
 TTGAACTTGTCAATGCGAAAGCGATATTCCAATATATACATGTAGTATTATTTATATGGA
 AGTTCTAAAAGGTG
 TTGAGTGGATTGTACTTGTGGAGGATGCTATTTCCTCTGCCATAATTTACGAGTA
 20 TGGATAACTACATA
 CTCATGTTATGAAACGCTCACTTATTGAAAACCTCTAACACACAAATATGCACTAGA
 TTCCAAACGCTAGACC
 AATTGTATCTAACTCAAATTCTCAATCAAAGTATTAATTACCGATGGTAAGAAAAGTTAAC
 GATATAATTATCAAA
 25 GAAAGAATAAGTCAACAGATTCTTAATCTTATTGGAACCTCTAACACACAAATATGCACTAGC
 TCAAAAGATATGTAAC
 TGTTAAAATAAATTCACTGAGATTAATTCTCAGACTCGTGTAGCTATAATAATGCAAG
 AGTTCTCTGTGTTCA
 AGGAAAAACCTTAAAGATATGTATTTCTGTAATTATGATGATATAATTGCTATTCTATTG
 30 CACAAACATTACTTAA
 AAAATCGTATTTCTACTACAAATGTTGACTAAGAACAAAATACATTGATTATTGATATAT
 CGTCAACTGAATTTC
 TTCCGAGGGATATAATTCTCAAACATAGCAAGAACATCTCATAATAATGTTCGTACTACCTTAA
 GACGAAATTTTTAA
 35 'GATTGTAACGTGACTTATGGTCTCTGCTGTGGGGTCAATGCGAATAATCTAAATGTATG
 GGAGTCAAATAAAATAC
 CAAGAAAAATAAAGGAGCAGCACCCAATAAACTATATGGGACCAGAAATCCTTCATTGGTT
 AAAATAGGATTATCCCG
 AAAGATGAAGGACTAAATTGAAACTGATTGGGGTAGGAAGAGATCCGTACAATCTTAAAT
 40 GGCTTCCACGCGGAAACT

TGTCGTTATACAATTCTTAACTTCGGGTCGGTTATATTCAAATGGGTCAAATAATAT
 TAGTTAACACTAA
 CGGAGTAATTAAATTGGTACTACAATTTCAGTTGGTCAATTAGAACGAAACATAGTCG
 TAAAATACGAGTCGGT

5 GITATACCTTATTCAGTIAAAAAAATACGAGATTITGTGTCAAATTCAAATTATTCAT
 GAATATATGGAAATTA
 TTAGATACTCTAGCGAAAATAGTGATTATGAGCGTTACAAAATACGATTITAGCATTGAA
 CTTCTTATGTAAATTC
 GGTCAAATGTTGGCATGAAGAAGCAAGTTGCAACATTAATTCAATTAAAATCGTGTGA
 10 CATACTTTAAAATCTAA
 ATATAGGAAGAAGACAAAACATTAATTAGTAAGATTCTAATGAACATTATAAGTTATAA
 CTTATAACCAACAAAAG
 TTGGGTTAGCGTTGTGCTTATCTGAAAACCTGCAAACCTAACCCATTAAATAGGACTAATG
 ACAATTAAACAACAAAAG

15 TACACTTAAGCAACAAGCCTCGTGAATATAATTGGGCCTCAGGCCATATTGCTAACGCC
 AACTGATAATTCACTT
 ATTCCCTCTTCATCTCACACACTCTCTCTATCTCTATCTCTAACGGCATAGCTGACTCAGT

20 SEQ ID NO:4 DMT 3' flanking sequence
 AGATGACTGGAAGAAAGCAACGCATGCTCTCTGCCTCTCTATTTAAAGCCAGGAAAAG
 TCCCATTTAGACATAAT
 AACAGGAATCCAATAGGCTATTCTCTTCTTCTTATTCATTCTATCTCTAACGGCATAGCTGAGCAGAAGCGACA
 CAAAAAAAGTTTTGG
 25 GTTATTTATTTCTCTCTAACAAAAAAAAAAAAAAACTCGAG

SEQ ID NO:5 DMT cDNA sequence

GTTCTCGGCATTGACTCGCTGAGAACATGAGAAAGCTAGATCGGTGAGCTTTAGCTCCATT
 TCTGTTATTTACATA

30 TTATTCCTTTTCTCTCCCTTTTATCTGGAATTGTTCTGCTAAATTTCAGCTGTTA
 CATTTCGGATCAC
 GAGAAGAACATCTGGGTTTATGTTAACATACATGTTCTGTTCTGATCATAAATCTCA
 GCTATTAAACACCTGAT
 TTGATTCGCGTAATAAAACCTCTGATTGCTTTATCTTCACTTCCCCATAAACATIGCTT
 35 ACTTTATTCGCTCTT
 CTTITACCGTTCCAGCTAAAAAATTCTCGCTATTCAATGTTCTGTTCTGTTGAGAA
 AAATATCTGACAAAAA
 AATCATTATTCGATTTATGGTGCAGATTCTAGTTAACATGTCGCCCTCTAACCAAGTCAGA
 TAAAAAAGGAGTGTTC

GTCCATGTTGCTTGGTGTGGAGAGAGTTTCGGAGAGTTAGGTGAGTGTATTGG
 GGTGAGGTAGTGATAA
 GGTTGAAGGGGGAGTGTATCAAGTGTGTTAATTGAGGGCTGATCCGGGGATAGA
 TATTTCGAGTTCCTT
 5 GGAGAACAACTCAACAAGAGTTCATGGGTTCTGGATTCCATTACACCCAAAAAACCTAG
 ATCAAGTGTGATGGTAG
 ATGAGAGAGTGTATAAACCCAGGATCTAAATGGGTTCCAGGTGGTAATTGTAGACAGGGGA
 TTCTGCAACACTGGTGTG
 GATCATAATGGGTTTGTATGGTGTATCAGGGCGTACCAACTAAGTATGATGATCA
 10 ATAGCTTAGCGGGATC
 ACATGCACAAGCTGGAGTAATAGTGAGAGAGATCTTGGCAGGAGTGAGGTGACTCTCC
 TTAGCACCACTTAC
 GAAACACCAACCGGTAATGTAGGCCGGTCAATGGAATTACTTCAGATGTGGTATGGTA
 ATGGTCCTTCACCCAG
 15 AGTGGCACTTCTCAAGCTGGCTATAATGAGTTGAATTGGATGACTTGTGAATCCTGATCAGA
 TGCCCTTCCCTCAC
 AAGCTTGTGAGTGGTGGGATAGCTTATTCAAGGTCGTCAATGTGAGTGATCAAATCTATT
 TCAGTTTTTTTTTC
 CCTTCTCCGTCTTCAGTACTTAGAGTAGAACATGAATTAGAATATCTAAGAAAGTCATG
 20 GTTTGAACAGATGGA
 CCTCCAGCGTGTAAACAAGCCTTTACAATTGAATTACCAATTAGAAGAGAACAGCTGG
 TCAGTCTGTGAAAGTT
 GTTCAATATGTACCGTCAACGCCAGTCTGTTAGAACAGGTGAAAAGACTGGATTCTIGA
 ACAGATAGTACAACTA
 25 CTGGACATGAAATCCCAGAGCCGAAATCTGACAAAGTATGCAGAGCATTATGGACTCGTCTG
 CTGTTAATGCGACCGAA
 GCTACTGAACAAAATGATGGCAGCAGACAAGATGTTGGAGTTGACCTAACAAACTCCT
 CAGCAGAAACCCCTCAA
 AAGAAAAGGAAGTTCATGCCAAGGGTGTGGAAGGCAAACCTAAAGAAAGGCCACGCA
 30 AACCTGCAGAACCTCCA
 AAGTGGTGTGGAAGGCAAACCTAAAAGGAAGGCCACGCAAAGCTGCAACTCAGGAAAAAGTG
 AAATCTAAAGAAACCGGG
 AGTCCAAAAAGAAAAATTGAAAGAACAGCAACTAAAAGGCCAGCCAATGTTGGAGATAT
 GAGCAACAAAAGCCCTGA
 35 AGTCACACTCAAAAGTTGCAGAAAAGCTTGAATTITGACTTGGAGAATCTGGAGATGCGAG
 GCAAGGGTACTCTGAGT
 CTGAAATTGTCAGAACAGTAGTGGCAGCAAACCTGTTCTGAGATCAGAGATGCCATTGGTG
 GAACATAATGGTAGTTG
 CTGGATTCACTGTGTCACAAATAGACAAGACCAATGGATTGGGGCTATGAACCAGCCACTTGAA
 40 GTGCAATGGGAAACCA

5 GCCAGATAAACTATCTACAGGAGCGAACTGGCCAGAGACCAACAACCTGATTATTGACTAG
 AAACCAGCAATGCCAGT
 TCCCACTGGCAACCCAGAACACCCAGTTCCAATGGAAAACCAACAAGCTTGGCTTCAGATGA
 AAAACCAACTTATGGC
 10 5 TTICCATTTGTAACCAGCAACCTCGCATGACCATAAGAAACCAGCAGCCTTGCTTGGCCATG
 GGTAAATCAACAACCTAT
 GTATCTGATAGGAACCTCACGGCCTGCATTAGTAAGTGGAAACCAGCAACTAGGAGGTCCCCA
 AGGAAACAAAGCGGCCCTA
 TATTTTGAAATCACCAAGACTTGTACCTGCTGGAAATCAGCTATATGGATCACCTACAGACAT
 15 10 GCATCAACTTGTATG
 TCAACCGAGGGCAACAAACATGGACTACTGATAAAAAACCAACCTGGATCATTAATAAG
 AGGCCAGCAGCCTTGCCT
 ACCTTGATTGACCAGCAACCTGCAACTCCAAAGGTTTACTCACTGAATCAGATGGTAGCT
 ACCAGCATGTICATCGC
 20 15 CTGGGCTTCGACCTCATCTCAGTCACAAGTTCTACAACATATCTACATGTGGAACTGTTTC
 CAGGATTTGAAATGGG
 ACTACAGGTACATGCCAGAGAAGCAGGGCTCTGCATACGATTCTTACAGCAAGATATCCAT
 CAAGGAATAAGTACAT
 ACTTTCTCATGAGATATCCAATGGAATGGGTGCAAGAAAGCGTTACCTCAAACCTCTCTG
 25 20 CCAACTCCAATTATGG
 CTAAAACTTGAGGAAGCCAGGGCTCGAAGAGACAGTATCATCGTCAATGGGACAGACGGAA
 AACGATGATCTAAACTTA
 GCTCAAGCAGATTGCTCAATCACAAGATGTGGAGAGACATAACAGCAGCACGTGTGGAATA
 TTAGATGCTGCAAAGAA
 30 25 AACGAAAATCCAGAAAGTAGTCCAAGAAAATTGATGGCATGCCACCTGAGGTTATAGAAA
 TCGAGGATGATCCAAC
 ATGGGGCAAGGAAAGGTTAAACTGCCAGCATCAGTAAAGGTGCATCTAAAGGAAACTCG
 TCTCAGTTAAAAGACA
 GCAGAAAAGGAGAAATGATTGTCACAAAGCCTGCAAAAAGGGTCAGCAGGTAGAAA
 AAAATCAGTACCTCCGCC
 35 30 TGTCATGCCAGAGATCAGCTTGGCAACCTACTCCCAAAGACACCTTATCAAGAAG
 CAAGCCTAAAGGAAAG
 GGAGAAAGTCATACAAGATTCAAGGAAAGCAAGAGGTCCATCAGGAGAACTCTGTGTCAG
 GATTCTATGCGGAAATA
 40 35 ATTTACAGGATGCAAATCTGTATCTAGGAGACAAAGAAAGAGAACAGAGCAAATGCAAT
 GGTCTTGTACAAAGGAGA
 TGGTGCACCTGTTCCCATGAGAGCAAGAAGCGAAAACCAAGACCCAAAGTTGACATTGACG
 ATGAAACAACTCGCATAT
 GGAACCTTACTGATGGGAAAGGAGATGAAAAGAAGGGATGAAGAGAAGGATAAAAGAA
 AGAGAAGTGGTGGGAAGAA

GAAAGAAGAGTCTCCGAGGAAGGGCTGATTCTTCATCGCTGCATGCACCTGGTACAAGGA
 GATAGACGTTTCGCA
 ATGGAAGGGATCGTGGTTGATTGGTACACAGAACATGTCGGATCA
 CCTTCAAGCTCTGCGT
 5 TCATGTCTCTAGCTGCTGATCCCTCCAAAATTAAAGCAGCAGCCGAGAAGATGAAAGGAATG
 TTAGAAGCGTAGTTGTT
 GAAGATCCAGAAGGATGCATTGAACTTAATGAAATTCTCGTGGCAGGAAAGGTTCAA
 CATCCATCTGACATGGA
 AGTTTCTGGGGTTGATAGTGGATCAAAGAGCAGCTAAGGGACTTCAAACCTCTGAAATTGA
 10 AAGATTTAATTCTTAG
 AGAAGAGTATTCAAATTAGAAGAGGAAGTATTATCATCACAAGATTCTTTGATCCGGCGA
 TATTCAGTCGTGTTGG
 AGAGTTGGATCCTGTTCATGTTCAAATCAGACGCAGAGTTCTACAACCAGGTGAAACA
 AAAACTGTCACTGGAAC
 15 ATCACAATCAGTGCACAACTGGAGCCAAACTTGTCTGATGAAATTGCTTCAAGGGAAATGA
 GAGACCGCATCTATATG
 AAGGATCTGGTGTGTTAGAAACAAGAAACTACAAATGTCGCTCAGAAGAAACCTGATCTTG
 AAAAAACAATGAATTGG
 20 AAAGACTCTGTCTGTTGGTCAGCCAAGAAATGATACTAATTGCAAACAACCTCTCCAGC
 AGCTATGAGCAGTGTGC
 GACTCGACAGCCACATGTAAGACATAGAGGATTTGGAATGCAAGGTGAAGGCCCTGGTTA
 TTCTTGGATGTCATCT
 CACCAAGAGTTGACAGAGTAAAGAACAAAAATGTACCCACGCAGGTTTCAGACAAGGTGGA
 25 AGTGTCCAAGAGAAC
 ACAGGTCAAGTCATACCATCAACGCCCATGAATTACCAAGGAATGGATTGTCGGTCTCA
 AGGCCGTCCAAGAAC
 CCAGGACGATACCAACATAATCAACAAGATGAGATGAATAAGCATCCCATTACAAAAAA
 CATTTTGGATCTGCTCA
 ACTCTCTGAAGAATGCCCTACAAGACAGTCAGTACCAACAGAACATCACGGATGGCTGTC
 30 TACCGAGAGATAGAACT
 GCTGAAGACGTGGTTGATCCGCTCAGTAACAATTCAAGCTTACAGAACATATTGGTCAATCA
 AATTCCACGAATAAAGA
 GCAGACGGCAGTTGAATACAAGGAGACAAATGCCACTATTTCAGAGAGATGAAAGGGACGC
 TTGCTGATGGGAAAAGC
 35 CTACAAGCCAGTGGGATAGTCTCAGAAAAGATGTGGAGGGGAATGAAGGGAGACAGGAACG
 AAACAAAAAACATATGGAT
 TCCATAGACTATGAAGCAATAAGACGTGCTAGTATCAGCGAGATTCAGGCTATCAAGGAA
 AGAGGGATGAATAACAT
 GTTGGCCGTACGAATTAAGGATTCTAGAACGGATAGTTAAAGATCATGGTGGTATCGACCT
 40 TGAATGGTTGAGAGAACAT

CTCCTCTGATAAAGCCAAGGACTATCTTGAGCATAAGAGGTCTGGTTGAAAAGTGTIG
 AATGCGTGCAGCTTA
 ACACCCACAATCTGCTTCCCTGTTGACACGAATGTTGAAGGATAGCAGTTAGGATGGGA
 TGGGTGCCCTACAAC
 5 CCTACCTGAAATCACTTCAGTTACACCTCTGGAGCTATAACCCAGTGCTCGAGTCCATCCAAAAA
 TTTCTTGCCAAAGAC
 TTGCAAACTCGATCAACGAACACTGTATGAATTACACTACCAACTGATTACGTTGGAAAGG
 TATTTGCAACAAGAGT
 AGACCAAATTGAAATGCAATGTCATGAGAGGAGAGTGAGACACTTGGCCAGTGCCTATGCT
 10 AGTGCAAAGCTTGCTT
 ACCGGCACAGAGGAGAGGAGCTAACAGTGCAACTATTCCGGTCCCTCCGAGTCCCTTCC
 TCCTGTAGGCCATCCCGA
 TGATAGAACTACCTCTCCGTTGGAGAAATCCCTAGCAAGTGGAGCACCATCGAATAGAGAAA
 ACTGTGAACCAATAATT
 15 GAAGAGCCGGCTCGCCCGGCAAGAGTGCACTGAAATAACCGAGAGTGATATTGAAGATGC
 TTACTACAATGAGGACCC
 TGACGAGATCCAACAATAAAACTCAACATTGAACAGTTGGAATGACTCTACGGGAACACAT
 GGAAAGAAACATGGAGC
 TCCAAGAAGGTGACATGTCAGGCTTGGTTGCTTGCATCCAACAACACTACTCTATTCCAAC
 20 TCCCAAACAAAGAAC
 ATTAGCCGCTCAGGACAGAGCACCAAGTGTACGAGCTCCAGATTCACATGTCCTTGTAT
 GGTATGGATAAAAGAGA
 ACCAGATGATCCAAGTCCTTATCTCTAGCTATATGGACACCAGGTGAAACAGCGAATTCCGC
 ACAACCGCTGAACAGA
 25 AGTGTGGAGGAAAGCGCTGGCAAAATGTGCTTGCAGAGACTGTTCTGAGTGTAAACAGTC
 TGAGGGAAGCAAACCTCA
 CAGACAGTCGAGGAACACTCTGTACCTTGCGACTGCCATGAGAGGAAGTTCCGCTC
 AACGGGACATATTTC
 AGTCAACAGTTATTCAGACCACGAGTCCAGTCTCAAACCCATCGATGTTCTAGAGATTG
 30 GATATGGATCTCCAA
 GAAGGACTTTACTTCGGAACATCAGTAACATCAATATTGAGGCTTCAACGGAGCAGA
 TACAGTTCTGTTTGG
 AAAGGATTGTTATGTCGGATTGAAACAGAAGACAAGAGCACCGCGTCCATTAAATGGCA
 AGGTTGCAATTCTCTGC
 35 GAGCAAATTGAAAGAACAAACAAACCTAAAGATGACTGGAAGAAAGCAAACGCATTGCTTCTC
 TGCTCTCTCTATTTAA
 GCCAGGAAAAGTCCCATTAGACATAAACAGGAATCCAAATAGGCTATTCTCTTCTTTC
 TTATITCATTCAAGA
 GCAGAAGCGACACAAAAAAGTTTTGGTTATTCTCTAACAAAAA
 40 AAAACTCGAG

SEQ ID NO:6 5' untranslated region of DMT

GTTCCTCGGCATTGACTCGCCTGAGAACATCAGAAAGCTTAGATCGGTGAGCTTTAGCTCCATT
TCTGTTTACATA

5 TTATTTCCCTTTCTCTCCCTTTATCTGGAATTGTTCTGCTAAATTTCAGCTGTTA
CATTTTCCGATCAC

GAGAAGAACATCGGGTTTTATGTTAACATACATGTTCTGTTCTGATCATAAATCTCA
GCTATTAAACACCTGAT

TTGATTCGCTAAACCTCTGATTGCTTTATCTCACITCCCCATAAACATTGCTT

10 ACTTTATTCGCTCTT
CTTTTACCGTTTCCAGCTAAAAAATTCTCGCTATTCAATGTGTTCTGTTGTTGAGAA
AAATATCTGACAAAA

AATCATTTATTGCTTTATGGTGAGATTCTAGTTAACATGTCGCCCTCTCAACCAAGTCAGA
TTAAAAGGAGTGTTC

15 GTCCATGTGCTTGGTGTGAGAGAGAGTTTCGGAGAGTTAGGTGAGTGTATTG
GGTAGGTTAGTGTATAA

GGTTGAAGGGGAGTGTATTCAAGTGTGTTGAATTGAGGGCTGATCCGGGGATAGA
TATTTCGAGTTCTT

GGAGAACAAACTCAACAAGAGTTCATGGGTTCTGGATTCCATTACACCCAAAAAACCTAG

20 ATCAAGTGTGATGGTAG
ATGAGAGAGTGTAAACCCAGGATCTAAATGGGTTCCAGGTGGTAATTGTAGACAGGGG
TTCTGCAACACTGGGTGT

GATCATAATGGGTTTTGATCATGGTGTCTCAGGGCGTACCAACTTAAGTATGATGATCA
ATAGCTTAGGGGATC

25 ACATGCACAAGCTGGAGTAATAGTGAGAGAGATCTTGGCAGGAGTGAGGTGACTTC
TTAGCACCACTGTTATCA

GAAACACCACCGTAATGTAGAGCCGGTAATGGAAATTACTTCAGATGTGGGTATGGTAA
ATGGTCCTTCAACCAG

AGTGGCACTCTCAAGCTGGCTATAATGAGTTGAATTGGATGACTTGTGAATCTGATCAGA

30 TGCCCTTCCTTCAC
AAGCTTGTGAGTGGTGGGATAGCTTATTCAAGGTTCTGCAATGTGAGTGATCAAATCTATT
TCAGTTTTTTTTTC

CCTTCTCCGTCTTGCACTTAGAGTAGAACATGAATTAGAATATCTAAGAAAGTCAG
GTTTGACAGATGGA

35 CCTCCAGCGTGAACAAGCCTCTTACAATTGAATTCCAATTAGAAGAGAACAGCTGG
TCAGTCTGTGAAAGTTTC

GTTCAATATGTACCGTAAACGCCAGTCTGTTCAGAACAGGTGAAAGACTGGATTCTG
ACAGATAGTTACAACTA

CTGGACATGAAATCCCAGAGCCGAAATCTGACAAAAGT

SEQ ID NO:7 >*Arabidopsis thaliana* DMT1 (1DMT5) gene sequence from BAC T32M21

(SAC-7406444):

73501 atcgcttcct tggatgggtgta agtcaatttt taactctc tatactcgag ttgtttcaact
 73561 tggacaacac tggtttaaaag tcctccatgg attaaaataac agatggataa aagagaacca
 73621 gatgatccaa tcgccttatct cttagctata tggcacccag gtgagaataa aactgcattg
 73681 ttccatcat gtgtctacag tttcaaaagaa agttagcgtt aagctaaaaaa gcatttgaaa
 5 73741 tagatcggtt taaaatgaa agtttgatc tggtaaatgaa agccggaaacg tagcatttgtt
 73801 ggatgttata tgtaaatttag tttttggatg tggctcaatg tagtgtttt actgcagggt
 73861 gaaaacagcga attccggcaca accgcctgaa cagaagtgtg gagggaaacg gtctggaaaa
 73921 atgtgccttgc accgacttgc ttctgagttt aacagtctgaa gggaaagcaaa ctcacagac
 73981 gttegaggaa ctcttcgtt gagatattct tggatctttt tggtgtctat gaaaaggaga
 10 74041 atgtggataa caatgttgc tttatcattt tttcgatcatt cacagatacc ttgtcgact
 74101 gccatgagag gaatgtttcc gtcacacggg acatatttcc aagtcaacacg ggtagatga
 74161 aataaaaactc aaacacacag acgaaacattt atttcgtttt atgtgtttt ctttatctc
 74221 ctggccattt ttatcttcg atgttatttgc agaccacagc tccagctca aaccatcoga
 74281 tggttcttaga gatggatataat gggatctccc aagaaggact gttaatccg gaaatcatgt
 15 74341 aacatcaata ttccggatgaaa aacatccg taatagatgtt agttaatcaa atgtccaaaa
 74401 cacaaggaaag cttaacccgtc caatacaca gaaagcttca cttttctttt gcaaaaaaaaa
 74461 atccatggatgtt gttttgtcag gtttttcaac ggagcagata cagtgttgc
 74521 tttggaaagg taaaacgttaa ctttccggcc agagaaatcc gggaaaatcta ttgtttttt
 74581 ctgtatcaata cggttttttttccat atacacacac accttttactt taggaccaat actgttctga
 20 74641 tctgtatgatgaa aacatcggtttaa acatcttaaca attatgttgc caggatttgtt atgtgtccgt
 74701 ggatgtcggac agaaagacaag accggccgtt ccattatgg caaggttgc tttttctcg
 74761 agcaaaatgtt agacaacaa aacctAAAGaa tgactgttgc aagcaacacg catgtttt
 74821 ctgtatcttctt ctatggaaatgg ccggggaaatgg tcccatatgg acataaaataac aggaatccaa
 74881 ataggttattt ttttttttttccatcaaaatgg ttttttttttccatcaaaatgg
 74941 ttttttttttccatcaaaatgg ttttttttttccatcaaaatgg
 25 75001 gtcacttagcg tggccaaatcc aatgtctcg cacaatcttggc cgcattgtca ataaaatttc

SEQ ID NO:8

ARABIDOPSIS THALIANA DMT2

>DMT2 (1DMT2) ;

MEKQRREESSFQQPPWI PQTPMKPFSPICPYTVEDQYHSSQLEERRFVGNGKDMGSLDHLS

5 FGDLLALANTASLIFSGQTPIPTRNTEVMQKGTEEVESLSSVSNNVAEQILKTPEPKPKR
 KHRPKVREAKPKREPCKPRAPIRKSVVTGQESKTPKRKVVRKKVEVKDQDATPVESSAA
 VETSTRPKRLCRRVLDFEAENGENTQNGDIREAGEMESALQEKGQLDSGNQELKDCLLSAP
 STPKRKRSQGKRKGVPKKNGSNLEEVDISMAQAACKRQGPTCCDMNLSGI QYDEQCDYQ
 KMHWLYSPNLQQGGMRYDAICSKVFGQHQHNYVSAFHATCYSSTSLSANRVLTVERRE
 10 GIFQGRQESELNVLSDKIDTPIKKKTTGHARFRNLSSMNKLVEVPEHLTSGYCSKPQQNN
 KILDTRVTVSKKPTKSEKSQTKQKNLLPNLCRFPPSFTGLSPDELWKRRNSIETISEL
 LRLLDINREHSETALVPTMNSQIVLFGGGAGAIVPVTVPKKPRPRPKVDDDETDRVWK
 LLLENINSEGVDGSDEQKAKWEEERNVFRGRADSFIARMHLVQGDRTFTPWKGSVDSV
 VGVFLTQNVDHLLSSAFMSLASQFPVPFPSSNFAGTSSMPSIQITYLDSEETMSSPP
 15 DHNHSSVTLKNTQPDEEKDYVPSNETSRSSSEIAISAHEVDKTTDSKEYVDSDRKGSVV
 EVDKTDEKCRVLNLFPSEDSALTQCQHSMVSDAPQNTERAGSSSEIDLEGERYRTSFMKLLQ
 GVQVSLLEDNSQVSPNMSPGDCSSEIKGQSMKEPTKSSVDSSEPGCCSQDGDVLSQCQP
 TLKEKGKVLKEEKKAFWDCLRREAQARAGIREKTRSTMVDWKAIRAADVKEVAETI
 KSRRGMNHKLAERIQYLTLNMKIMQGFLDRLVNDHGSIDLEWLRDVPPDKAKEYLLSFNGL
 20 GLKSVECVRLLTLLHHLAFTPDTNVGRIA VRLGWVPLQPLPESLQLHLLEMYPMLESIQKY
 LWPRLCQLDQKTLYELHYQMITFGKVFTKSKPNCNACPMKGECRHFAFARKFSNIHL
 FY SARLALPSTEKGGMGTPDKNPLPLHLPPEPFQREQGSEVVQHSEPAKKVTCCEPIIEEPA
 SPEPETAEVSIADIEEAAFFDPEEIPITIRLNMDAFTSNLKKIMEHNKELQDGNMSSALVA
 LTAETASLPMPLKKNISQLRTEHRVYELPDEHPLLAQLEKREPDDPCSYLLAIWTPGETA
 25 DSIQPSVSTC1FQANGMLCDEETCFSCNSIKETRSQIVRGTTI1PCRTAMRGSFPLNGTY
 FQVNEVFADHASSLNPINVPRELIWELPRRTVYFGTSVPTIFKGLSTEKIQACFWKGYVC
 VRGFDRKTRGPKPLIARLHFPA SKLKGQQANLA

SEQ ID NO:9

>DMT2 (1DMT2) novel 480 amino acid amino terminus;

MEKQRREESSFQQPPWI PQTPMKPFSPICPYTVEDQYHSSQLEERRFVGNGKDMGSLDHLS

30 FGDLLALANTASLIFSGQTPIPTRNTEVMQKGTEEVESLSSVSNNVAEQILKTPEPKPKR
 KHRPKVREAKPKREPCKPRAPIRKSVVTGQESKTPKRKVVRKKVEVKDQDATPVESSAA
 VETSTRPKRLCRRVLDFEAENGENTQNGDIREAGEMESALQEKGQLDSGNQELKDCLLSAP
 STPKRKRSQGKRKGVPKKNGSNLEEVDISMAQAACKRQGPTCCDMNLSGI QYDEQCDYQ

KMHWLYSPNLQQGGMRYDAICSKVFSQQHNYVSAFHATCYSSTSSQLSANRVLTVEERRE
 GIFTQGRQESELNVLSDKIDTPPIKKKTTGHARFRNLSMNKLVEVPEHLTSGYCSKPQQNN
 KILVDTRTVSKKKPTKSEKSQTQKQNLLPNLRCRFFPSFTGLSPDELWKRRNSIETISEL

5 SEQ ID NO:10

>DMT2 (1DMT2) Nucleotide sequence from BAC F1011 (gi_6598632),
 60001 tgctcgagcc tgggttttctt catcgaccc ggatcttcgg atcttatcaa cggctcaaga
 60061 tcgttgtccaa ttccaaagaa ctatacataa caaggatgc gcaataatgc cccataaatta
 60121 aacctaatcg goaaaaatcg attcgactg caaacaatcc tcgttagagg ggaatttcaga
 60181 gcattacac aatacgtaaa cttaaatgtac aatctaaaaa ttgagatgoa taacgcggatt
 60241 ctgcgaaagaa gacggagaag atagaaggaa tgcttcgaat tggcaaaaaa tgcagagag
 60301 ttggcaatc ctgcgatcc tttaggtgtb gaattggggg tttttatggg acggacaaaa
 60361 aaaaatgttga agatcgagc tggttccaaa aatatttgg cccatataat gaccacatt
 60421 ccatgtataac tagggccatc actataatc tgacaaatc agaattctt ggtcggttg
 60481 aactatctga tttaaaaaacaa gttaaatggag atccctccaca tatacgaaacca gatcttgatt
 60541 caggtaaccc aaatgtcaacc gttaaaatccg atataaaacca aacggaaaggaa acagagagtt
 60601 tacacatcg ctggctgttcc ttgtgtgaca aatgttttgcgatc aatggaaacta
 60661 aaatggatc tagaaaaatccgttataacttggatcggatcggatcggatcggatcggatcggatc
 60721 tacattaact agatggaaac agtgtgtctt ctcaacttta accataagct tgatatttt
 60781 tgctgtccaa cggatTAggc gagggtatgc tggtgtccct tcgtttgtc cgccggaaag
 60841 tgcataatcg cttaaaatccg ctggctgttcc tcgtgttccgatc aatggaaacta
 60901 acgttccatcataataatca aaaaatgtat tgatgttccg aacaaagaa aactgttttca
 60961 tacaatgttccaa aagcaagctt gtatcttcg atgttgcataa cctggaaaaaa
 61021 atagatcgcc agtgtgtttaga aacatctttt ttatataaaa caatcgatc acatccaaat
 61081 aaagaaaaatccatccatccatggatatcgta ggaacagaggg ttttttttttttttttttttttt
 61141 cggatgttccatccatccatccatccatccatccatccatccatccatccatccatccatccatcc
 61201 tccggcaaaatccatccatccatccatccatccatccatccatccatccatccatccatccatccat
 61261 aagagtgttga gtattttggat tcgttgcatttccatccatccatccatccatccatccatccat
 61321 agggaaaaac tacccctcatc ogctgttccatccatccatccatccatccatccatccatccat
 61381 atctgttagaa aacataatcg tagtaaagccg ctttttttttttttttttttttttttttttttt
 61441 aattgttccccccatccatccatccatccatccatccatccatccatccatccatccatccatccat
 61501 ctccctgtccaaatccatccatccatccatccatccatccatccatccatccatccatccatccat
 61561 aagatccatcgccgttccatccatccatccatccatccatccatccatccatccatccatccatcc
 61621 gatt
 61681 atggatccatccatccatccatccatccatccatccatccatccatccatccatccatccatccat
 61741 ttttccatccatccatccatccatccatccatccatccatccatccatccatccatccatccatcc
 61801 ttggatgttccatccatccatccatccatccatccatccatccatccatccatccatccatccatcc
 61861 attt
 61921 gtt
 61981 ttt
 62041 tccgtgttccatccatccatccatccatccatccatccatccatccatccatccatccatccatcc
 62101 ttccatccatccatccatccatccatccatccatccatccatccatccatccatccatccatccat
 62161 ttccatccatccatccatccatccatccatccatccatccatccatccatccatccatccatccat
 62221 ttccatccatccatccatccatccatccatccatccatccatccatccatccatccatccatccat
 62281 ttccatccatccatccatccatccatccatccatccatccatccatccatccatccatccatccat
 62341 ttccatccatccatccatccatccatccatccatccatccatccatccatccatccatccatccat
 62401 agggaaaaatccatccatccatccatccatccatccatccatccatccatccatccatccatccat
 62461 tgcactgtttaa aacaaatgttga ttgttactaaa tttttttttttttttttttttttttttt
 62521 tagatccatccatccatccatccatccatccatccatccatccatccatccatccatccatccat

卷之三

65641 tgccatgtt gttttttttt tttttttttt ttttttttttt ttttttttttt
65701 ggaaatatac accaaatgtat ttttttttttt ttttttttttt ttttttttttt
65761 agatctatgtt ccgacaaatcg tggttttttt ttttttttttt ttttttttttt
65821 ctgcattttt ctatattttt cggattttt ttggttttttt ttttttttttt
65881 ctggcggtt tttttttttt ttgggggggg gttttttttt ttttttttttt
65941 ccgaatctt actgtttttt ttgggggggg gttttttttt ttttttttttt
66001 cgggtttttt ttgggggggg gttttttttt ttgggggggg gttttttttt
66061 tttagggacgg ttgggggggg gttttttttt ttgggggggg gttttttttt
66121 ttacttgcattt caactttttt ccggccattt ttccctttttt ttgggggggg
66181 tcgggtttttt ttgggggggg gttttttttt ttgggggggg gttttttttt
66241 ctacggaccc ttgggggggg gttttttttt ttgggggggg gttttttttt
66301 tcacggacat ttgggggggg gttttttttt ttgggggggg gttttttttt
66361 acctttttttt ttgggggggg gttttttttt ttgggggggg gttttttttt
66421 aatttgacca aacggattttt ccggccattt ttccctttttt ttgggggggg
66481 ggtttttttt ttgggggggg gttttttttt ttgggggggg gttttttttt
66541 cccaaaaaaa aatggatccat ttccctttttt ttgggggggg gttttttttt
66601 ttcaatccat ttccatccat ttccatccat ttccatccat ttccatccat
66661 aaaaaaaacccaa aaccccttttcc ttccatccat ttccatccat ttccatccat
66721 ggcagatccat ttccatccat ttccatccat ttccatccat ttccatccat
66781 ttgtttttttt ttccctttttt ttccatccat ttccatccat ttccatccat
66841 actttttttt ttccatccat ttccatccat ttccatccat ttccatccat
66901 gttttttttt ttccatccat ttccatccat ttccatccat ttccatccat
66961 aaatgtatccat ttccatccat ttccatccat ttccatccat ttccatccat
67021 gatgtttttttt ttccatccat ttccatccat ttccatccat ttccatccat
67081 tttagatccat ttccatccat ttccatccat ttccatccat ttccatccat
67141 ttccatccat ttccatccat ttccatccat ttccatccat ttccatccat
67201 tcactgtttttt ttccatccat ttccatccat ttccatccat ttccatccat
67261 ctttttttttt ttccatccat ttccatccat ttccatccat ttccatccat
67321 ttccatccat ttccatccat ttccatccat ttccatccat ttccatccat
67381 cggatccat ttccatccat ttccatccat ttccatccat ttccatccat
67441 atgatccat ttccatccat ttccatccat ttccatccat ttccatccat
67501 gtatccat ttccatccat ttccatccat ttccatccat ttccatccat
67561 ttatccat ttccatccat ttccatccat ttccatccat ttccatccat
67621 atatccat ttccatccat ttccatccat ttccatccat ttccatccat
67681 acatccat ttccatccat ttccatccat ttccatccat ttccatccat
67741 acgttatccat ttccatccat ttccatccat ttccatccat ttccatccat
67801 tagatccat ttccatccat ttccatccat ttccatccat ttccatccat
67861 tgccatccat ttccatccat ttccatccat ttccatccat ttccatccat
67921 caaatccat ttccatccat ttccatccat ttccatccat ttccatccat
67981 tgccatccat ttccatccat ttccatccat ttccatccat ttccatccat
68041 gtatccat ttccatccat ttccatccat ttccatccat ttccatccat
68101 ttatccat ttccatccat ttccatccat ttccatccat ttccatccat
68161 gttttttttt ttccatccat ttccatccat ttccatccat ttccatccat
68221 aaaaaaaatccat ttccatccat ttccatccat ttccatccat ttccatccat
68281 agtggatccat ttccatccat ttccatccat ttccatccat ttccatccat
68341 caactttttt ttccatccat ttccatccat ttccatccat ttccatccat
68401 acgttatccat ttccatccat ttccatccat ttccatccat ttccatccat
68461 aagtttttttt ttccatccat ttccatccat ttccatccat ttccatccat
68521 tgaatccat ttccatccat ttccatccat ttccatccat ttccatccat
68581 aagtggatccat ttccatccat ttccatccat ttccatccat ttccatccat
68641 tgaatccat ttccatccat ttccatccat ttccatccat ttccatccat

68701 aaatatatga gttggaaagtc atttttggac tatttatataa gatctaatta tcacacgacg
68761 ttggatgtt tggttagc agttgttgc agagagttcg ataaggccat cactccaaac
68821 atacaaaata tccatacatt gatccaccaa tataaccggc tgtgtgccaa gcaaagtgaa

5 SEQ ID NO:11

ARABIDOPSIS THALIANA DMT3

>DMT3 (1DMT3) ;

MEVEGEVREKEARVKGRQPETEVLHGLPQEQSIFNNMQHNHQPDSDRRRLSLENLPGLYN
MSCTQLLALANATVATGSSIGASSSSLSSQHPTDSWINSWKMDSNPWTLSKMQKQQYDVS
10 TPQKFLCDLNLTPEELVSTSTQRTEPESPQITLKTPGKSLSETDHEPHDRIKKSVLGTGS
PAAVKKRKIARNDEKSQLETPTLKRKKIRPKVVREGTKKASSKAGIKKSSIAATATKTS
EESNYVPRPKRLTRRSIRFDLQEEDEFCGIDFTSAGHVEGSSGEENLTDTTLGMFGHV
PKGRGRGQRRSNGFKKTNDCLSSMLSLVNTPGPSFMESEEDRPSDSQISLGRQRSIMATR
PRNFRSLKLLQLQRIIPSKRDRKGCKLPRGLPKLTVASKLQLKVFRKKRSQRNRVASQFNA
15 RILDLQWRRQNPNTGTSLADIWERSLTIDAITLEKLFEELDINKEGLCLPHNRETALILYKK
YEEQKAIIVKYSKKQKPKVQLDPETSRVWKLLMSSIDCDGVGSDEEKRKWEEERNMFHG
RANSFIARMRVVQGNRTFPWKGSVVDSSVVGFLTNQVADHSSSSAYMDLAAEFPVENF
NKGSCHEWGSSVTQETILNLDPRTGVSTPRIRNPTRVIEEIIDDDENDIDAVCSQESSK
TSDSSITSADQSXTMLLDPFNTVLMNEQVDSQMVKGKGHIPYTDDLNLSQLGISMVSSAS
20 THCELNNEVPPEVELCSHQDPESTIQTQDQESTRTEDVKKNRKKPTTSKPKKKSSES
AKSTQKKSVDWDSLKEAESGGRKERTERTMDTVDWDALRCTDVHKIANIIKRGMNMM
LAERIKAFLNRLVKKHGSIDLEWLDRVPPDKAKEYLLSINGLGLKSVECVRLLSLHQIAF
PVDTNVGRIAVRLGWVPLQPLPDELQMHLLELYPVLESVQKYLWPRLCLKLDQKTLYELHY
HMITFGKVFCTKVKPNCNACPMKAECRHYSARASARLALPEPESDRTSVMIHERRSKR
25 KPVVVNFNRPSSLFLYQEKEQEAQRSQNCEPIIIEPPASPEPEYIEHDIEDYPRDKNNVGTSE
DPWENKDVIPTIILNKEAGTSHDLVNVKEAGTSHDLVVLSTYAAAIPRKLKIKEKLRT
HHVPELPDHHSILEGFERRERAEDIVPYLLAIWTGETVNSIQPPKQRCALFESNNNTLCNE
NKCFQCNKTREEESQTVRGTTILIPCRTAMRGGFPLNGTYFQTNEVFADHDSSINPIDVPT
ELIWDLKRRVAYLGSSVSSICKGLSVEAIKYNFQEGLYVCVRGFDRENRKPKSLVKRLHCS
30 HVAIRTKEKTEE

SEQ ID NO:12

>DMT3 (1DMT3) novel 375 amino acid amino terminus;

MEVEGEVREKEARVKGRQPETEVLHGLPQEQSIFNNMQHNHQPDSDRRRLSLENLPGLYN
35 MSCTQLLALANATVATGSSIGASSSSLSSQHPTDSWINSWKMDSNPWTLSKMQKQQYDVS
TPQKFLCDLNLTPEELVSTSTQRTEPESPQITLKTPGKSLSETDHEPHDRIKKSVLGTGS

PAAVKKRKIARNDEKSQLETPTLKRKKIRPKVVREGKTKASSKAGIKKSSIAATATKTS
 EESNYVRPKRLRRSIRFDLQEEDEEFCGIDFTSAGHVEGSSGEENLTDTLGMFGHV
 PKGRRGQRRSNGFKKTNDNDCCLSSMLSLVNTGPGSFMESEEDRPSDSQISLGRQRSIMATR
 PRNFRSLKKLLQRII

5

SEQ ID NO:13

>DMT3 (1DMT3) nucleotide sequence from BAC T22K18 (gi 12408726);
 53341 atccaatgtac taatcgacat ttaagggggg tgatttgcg gggttaaaac ggtttccaa
 53401 cggaaatcgtc ctgttccaaa ctgttacca ttgttttacc ccgtttgttc atgggttcc
 53461 tctttttccat ttggccggaa cactcaatga cacttttcc tccttttgc ttggccggac
 53521 accagaatgc agcatttaat ctctcaaaatt accattttgc tccttacatc tttaaccctt
 53581 ttgtttttttt ttgtttttttt ttgtttttttt ttgtttttttt ttgtttttttt
 53641 tatttttttccatgc acattttttttt ttgtttttttt ttgtttttttt ttgtttttttt
 53701 ttataaaaga aaatcatgtt taaacggaca ttgtttttttt ttgtttttttt ttgtttttttt
 53761 agttataaga gagataataa gatttttca cagagactca cttagtggaa aaaggaaagaa
 53821 gcaagggtttt cccttaagggg aaaaaaaaaaaggtt cactatcacat tttaatcgtt
 53881 actgtgtaaat tcaatagatc gtgcgttattt attcgttgcgtt attgttgcgtt aagaaatttat
 53941 ttgttcaatgtt taaatcaaggcc tcctttttttt ttgtttttttt ttgtttttttt
 54001 agtgtttttttt ttgttgcgtt aacacttcgtt tttaattttttt ggtttttttt acattttcc
 54061 acgggtttttttt ttgttggggcc aaagaagccaa cttagttttttt tagttttttt tagttttttt
 54121 ttgtttttttt ttgttccatc ttgttccatc ttgttccatc ttgttccatc ttgttccatc
 54181 cactgttgc caacatcatat ttgttccatc ttgttccatc ttgttccatc ttgttccatc
 54241 ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 54301 ttataatctt aattttttttt ttatccatc ttatccatc ttatccatc ttatccatc
 54361 caatgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 54421 cttagaagctg attttttttt aattttttttt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 54481 attgtatataa ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 54541 atatccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 54601 tttaggtttttt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 54661 attaagatgtt aattatggaa ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 54721 cattttttttt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 54781 ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 54841 ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 54901 ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 54961 ctccatccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 55021 gacttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 55081 caatgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 55141 attttttttt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 55201 accttccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 55261 acaatccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 55321 atgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 55381 ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 55441 ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 55501 ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 55561 ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 55621 aactgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 55681 ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc
 55741 aactgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt ttgttgcgtt
 55801 atgttgcgtt ttatccatc ttatccatc ttatccatc ttatccatc ttatccatc

58921 agcataaaacg gattaggatt gaagagtgtg gagtgtgtta gactttgtc actacatcg
 58981 attgcattcc ctgttaaagtca atgaaggata ctgaatactc agaccctaata gaatgtggaa
 59041 cagatacatt aatagttaag tatttttaca aatgcagggt gacacgaaatc tggacggcat
 59101 acgtgttaaga cttagatggg ttccccatca gcoaatcoco gacgagctgc aaatgcattct
 59161 ttttagatgg taagaaaaaaa aatattaaga tcattttca atcatgaaa ggaacatgag
 59221 aaattttacag tagtccctt taattttcat caggtaccca gtcttagatc cagtcataaa
 59281 gtacccctgg ccacgcctct gcaagctgtc cccaaaaacc ttgttaatgg attacatgg
 59341 catcaacatc tctttttatc cttttttatc ctctttatc ctctttatc atccatataat
 59401 ttttagatgg acgtgttca ccacatgttca acatgtggaa aggtttatc aacaaatttc
 59461 aagtgtttgtt ggtttttttt aatctttttt tggtttttttt tttttttttt gttttttttt
 59521 caaaaatggaa acccaatttc aatgcatttc caatgttca ggtttttttt gttttttttt
 59581 gtgcacgtgg aagggtttttt cccaaaaatc ttttttttttt tgccatataa atggaaaaaaa
 59641 aacatgttca tttttttttt aatctttttt tggtttttttt tttttttttt tttttttttt
 59701 cttaatgttcc ttgtttttttt gacggatgg gttttttttt ccacacccgg aggaggatgt
 59761 cagaacaaatgtt gtaatgttcc atggaggagg atctttttttt aacccatgttcc tttttttttt
 59821 tttttttttt tttttttttt atctttttttt aatctttttt tggtttttttt tttttttttt
 59881 tttttttttt tttttttttt aatctttttt tggtttttttt tttttttttt tttttttttt
 59941 acgttatcttcc cggggatggaa aacacgtttcc aacatgttcc gttttttttt tttttttttt
 60001 ctgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60061 ctggggatgtt gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60121 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60181 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60241 atccatatactt aatctttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60301 ggggtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60361 aactgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60421 atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60481 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60541 actaatgttcc aacaaatgttcc tttttttttt tttttttttt tttttttttt tttttttttt
 60601 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60661 atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60721 atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60781 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60841 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60901 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 60961 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61021 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61081 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61141 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 611501 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 611561 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61321 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61381 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61441 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61501 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61561 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61621 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61681 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61741 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61801 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61861 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
 61921 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt

61981 aatctctgat tcagattcaa gaaagacaaa gcatgagaca tcattctgca agttaaccgg
 62041 tccgggttat tctcgaaatcc tactgaatta agcataatc atctaaagga acttcataag

5 SEQ ID NO:14

Arabidopsis thaliana DMT4

>DMT4 (1DMT4) ;

MEFSIDRDKNLLMVPETRIKTKQFEKVYVRRKSIKLPQNSMVHNTLIKMARQRIQKSMK
 ESVMNQHIFKNFDSYLSVIYHPCCFVINNSQTTHKKKEKKNSKEHGIKHSESEHLQDDI
 10 SQRVTGKGRRRNSKGTPKKLRFNRPRILEDGKKPRNPATTRLRTISNKRRKKDIDSEDEV
 IPELATPTKESFPKRRKNEKIKRSVARTLNFQKEIVLSCLEFDKICGPIFPRGKKRRTTR
 RRYDFLCFLLPMPVWKQSRRSKRKNMVRWARIASSSKLLEETLPLIVSHPTINGQADA
 SLHIDDTLVRHVVSQTKKSANNVIEHNRQITYQKDGLSSLADVPLHIEDTLIKSASS
 VLSPERPIKKTDIAKLIKDMGRLKINKKVTITMIAKDKKLVTAKVNLDPETIKEWDVLMVN
 15 DSPRSYDDKETEAKWKKEREIFQTRIDLFINRMHRLQGNRKFKQWKGSSVVDVGVFLT
 QNTTDYLSNAFMVAAKFPVDAREGLSYYIEEPQDAKSSCIIILSDESISKVEDHENTA
 KRKNEKTGIIIEDEIVDWNNLRRMYTKEGSRPEMHMDSVNWSDVRLSGQNVLLETTIKRGQ
 FRILSERILKFLNDEVNQNGNIDLEWLNRNAPSHLVKRYLLEIEGIGLKSACVRLGLKH
 HAHPVDTNVGRIA VRLGLVPLEPLPVGQVMHQLFEYPSMSDIQKYLWPRLCKLPQETLYE
 20 LHYQMITFGKVFCTKTIPNCNACPMKSECKYFASAYVSSKVLLESPEEKHMHEPNTFMNAH
 SQDVAVDMTSNINLVEECVSSGCSDQAIKYKPLVEFPSSPRAEIPESTDIEDVPMNLYQ
 SYASVPKIDFDLALKSVEDALVISGRMSSDEEISKA LVIPTPENACIPIKPPRKMKY
 YNRLRTEHVVYVLPDNHELLHDFERKLDDPSPYLLAIWQPGETSSSFVPPKKKCSSDGS
 KLCKIKNCSYCWТИREQNSNIFRGTTILVFAHETSLNPIVFRRELCKGLEKRALYCGSTV
 25 TSIFKLLDTRRIELCFWTGFLCLRAFDRKQRDPKELVRRLHTPPDERGPNGFHIVVVDEK
 EESPRVGLMVMMPGFWGGSVIQNRVYVSGVKVLE

SEQ ID NO:15

>DMT4 novel 372 amino acid NH2 terminus;

MEFSIDRDKNLLMVPETRIKTKQFEKVYVRRKSIKLPQNSMVHNTLIKMARQRIQKSMK
 ESVMNQHIFKNFDSYLSVIYHPCCFVINNSQTTHKKKEKKNSKEHGIKHSESEHLQDDI
 SQRVTGKGRRRNSKGTPKKLRFNRPRILEDGKKPRNPATTRLRTISNKRRKKDIDSEDEV
 IPELATPTKESFPKRRKNEKIKRSVARTLNFQKEIVLSCLEFDKICGPIFPRGKKRRTTR
 RRYDFLCFLLPMPVWKQSRRSKRKNMVRWARIASSSKLLEETLPLIVSHPTINGQADA

SLHIDDTLVRHVVKQTKKSANNVIEHLNRQITYQKDHGSSLADVPLHIEDTLIKSASS
VLSPERPIKKT

SEQ ID NO:16

5 >DMT4 nucleotide sequence BAC F28A23 (gi 7228244);
14881 gctatggatc tcaacagaga gaattacgaa ttgggtttac cgatcatgg aaaaaggccgc
14941 gttgctcaca agatcgacctt cggggaaaggc cttgtcttc cogttcttg tgaatacggt
15001 gctgcgactaa gcattttttt ttctgcgactt attaacaaaaa aagatgtgc agataatggaa
15061 attaaaaaaa actatggcc taattttttt gatataatatt ctgtatggaa attttctgtat
15121 acggcaatttt gttttttt tcaacagaga aagtgtgtt actgttattt ttgtgtgttt
15181 gggtgcggaa gaaggaaatcc ggaacatattt acctttatattt cttgtatgtt gacaagacaa
15241 actacatcaa ctaccacaaag cttttgtatcg atctttgtgaa aattttggggaa gtgatttttt
15301 acggcaacac tctgtggaaat gttttgtctgg tggctcttcc tttttttttt gatggaaatgg
15361 aegttgttca ctacacagac ttgtttttttt gatggaaacaa ggctttgtt gctgaccctc
15421 ggatcgatgtt ctgtatgttcc ttgtttttttt atggaaatcc tttttttttt tttttttttt
15481 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
15541 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
15601 atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
15661 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
20 15721 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
15781 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
15841 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
15901 atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
15961 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
25 16021 aactttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16081 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16141 aactttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16201 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16261 aactttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
30 16321 aactttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16381 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16441 aactttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16501 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16561 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
35 16621 aactttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16681 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16741 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16801 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16861 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
40 16921 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
16981 aactttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
17041 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
17101 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
17161 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
45 17221 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
17281 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
17341 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
17401 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
17461 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
50 17521 agatgtttca caacacacccat atcaaaaaatgg cgagacaaag aatcaaaaaag agatgtttca

17581 aaagtgtat gaatccacat atctttcaagg taaaattttaattt cttaaaaaaag
17641 ttatgcctttt ggtaatggta ttacaattta tatttttttttt acggatgtt
17701 agtttgataa gatccatcca aacccgttcc tccatcgcc atcgatgtt
17761 atcccttgcattt ttatgttgcattt agggatttt cagaatttcc attcatcatt
17821 tttatccatc gtgtttttt aatcgtttaa ttcaacttattt ttatttttttt
17881 aacagataat tcacaacca cccataaaaaa aaaggagaag aagaatccaa aaaaaaaagca
17941 tggaaataaag catctgcattt cagaatccatc tcgaatggaa tacttttgcattt
18001 aaaaaaaaaacag ttatgttgcattt aggttataattt tttttttttt
18061 atgtttttt gcaacatcca ctaatcattt cccacttcattt cattttatattt
18121 gtgaaatggat atatttgcac acgtgttacc ggaaaaaggaa ggagaaggaa ttcaaaagggg
18181 acacccaaaaaa aacccgggtt taatggccat cggatccgtt aacggccaa gaaaaacca
18241 aatcccggccaa cccatctggat gagaatccatc tccaaacaacca ggaggaaaaa ggacatag
18301 agtggaaatggat aagtttattttt gaaatggccatc actccaaacca agggaaatgg
18361 aaaaaaaacgg agaaagattaa gagatccgtt gtcggacattt taaaattttaa
18421 gttctggatgtt gtcggatgtt gcaacatcca ttccatccatc
18481 aggaccacca caacggccat atatggccat cttttttttt tacttccatg
18541 aaaaaaaacatc caaaggatccatc taatggccat cttttttttt tacttccatg
18601 ttctttccatc actgtgttca agaaatccatc cttttttttt tacttccatc
18661 ggacaaggccat atgttttttccatc acacatgtt gttatccatc
18721 ctgttttttttccatc aaaaatggat atgttttttccatc
18781 gacacatccatc tgatccatc atgttttttccatc
18841 gacccatccatc atccatccatc aaaaatggat gtttttttttccatc
18901 gtttttttttccatc acatccatc
18961 aaaaaatggat ttttttttttccatc
19021 ttttttttttccatc
19081 accatccatc aaaaatggat attttttttccatc
19141 caataaaaaaa gtaacaacca tgatccatc
19201 ttttttttttccatc
19261 atccatccatc
19321 cccatccatc
19381 attttttttccatc
19441 ggaatccatc
19501 ttttttttttccatc
19561 atgttttttccatc
19621 gtttttttttccatc
19681 attttttttccatc
19741 gatccatccatc
19801 ttttttttttccatc
19861 agggaaaaacgg
19921 aggttgcata
19981 gacgttgcata
20041 aggttttttccatc
20101 aattttttttccatc
20161 ttcttttttccatc
20221 catccatccatc
20281 gtaatggccat
20341 gatggccat
20401 ggttttttttccatc
20461 accatccatc
20521 ttttttttttccatc
20581 ttttttttttccatc

SEQ ID NO:17

RICE(*Oryza sativa*) DMT1**>DMTRICE (1DMTRICE) ;**

MQDFGQWLPQSQTADLYFSSIPSPQFDTSIETQTRTSAVSSEKES

5 ANSFVPHNGTGLVERISNDAGLTEVGSSAGPTECIDLNLTPARKPKKKHRPKV
 LKDDPKSKTPKSATPIPSTEKVEKPSGKRKYVRKKTSPGQPAAEQAASSHCRSELK
 SVKRSLDFFGEVLQESTQSGSQVPVAEICTGPKRQSIPTIQRDSQLACHVVSST
 SSIHTSASQMVNAHLFPDNMPNGVLLDNNSTSSQLQNEHAKFVDSPARLFGSRI
 RQTSGKNSLLEIYAGMSDRNVPDLNSSISQTHSMSTDFAQYLLSSSQASVRETQM

10 ANQMLNGHMRMPENPITPSHCIERAALKEHNLHVPHAKAAVMNGQMPHSYRLAQ
 NPILPPNHIEGYQVMENLSELVTNDYLTTASPFSQTGAANRQHNIGDSMHIALD
 PRRESNASSGSWSISLGVNFNQQNNGWASAGAADAASSHAPYFSEPHKRMRTAYL
 NNYPNGVVGHFSTSSTDLSNNENENVASAINSNTLADAQRILIAREKSRSQRM
 ISFRSSKNDMVNRSEMVKHQHGRPAFHGSACRESIEVPDKQFGLMTEELTQLPSMP

15 NNPQREKYIPQTGSCQLQSLEHDMVKGHNLAGEHLHKQVTSPQVVIQSNFCVTTP
 DVLRRTSGEHLRTLIAPTHASTCKDTLKALSCQLESSRDIIRPPVNPIGPSSADVP
 RTDNHQVKVSEETVTAKLPEKRKVRPRKELKPGEKPKPRGRPRKGKVVGELA
 SKDSHTNPLQNESTSCSYGPYAGEASVGRAVAKANRVGENISGAMVSLLSDLSIVI
 QKIKVLDINKSEDPTAEPHGALVPYNGEFGPVPFEGKVKRKRSRAKVLDLDPV

20 ALMWKLLMGPDMSDCAEGMDKDEKEWLNEERKIFQGRVDSFIARMHVLQGDR
 RFSPWKGSVVDSVGVFLTQNVDHLSAFMALAAKFPVKPEASEKPNAMVFH
 TISENGDCSGLFGNSVQLQGEILVQEASNTAASFITTEDKEGSNSVELLGSSFGDG
 VDGAAGVYSNIYENLPARLHATTRPVVQTGNAVEAEDGSLEGVVSSENSTISSQN
 SSDYLFHMSDHMFSSMLLNFTAEDIGSRNMPKATRTTYTELLRMQUELKKNKSNETI

25 ESSEYHGVPVSCSNNIQVLNGIQNIGSKHQPLHSSISYHQTGVHLPDIVHASDLE
 QSVTGLNRVLDNSVTQTSYYPSPHPGIACNNETQKADSLSNMLYGIDRSDKTTS
 LSEPTPRIDNCFQPLSSEKMSFAREQSSSENLYSRNEAAFKQHGTNSVQGDN
 TVRTEQNGGENSQSGYSQQDDNVGFQTATTSNLYSSNLCQNKQKANSEVLHGVS
 NLIENSKDDKTPKVPVDGSKAKRPRVGAGKKTYDWDMRKEVLYSHGNKE

30 RSQNAKDSIDWETIRQAEVKEISDITRERGMNNMLAERIKDFLNRLVRDHGSIDLE
 WLRYVDSDKAKDYLSSIRGLGLKSVECVRLTLHHMAFPVDTNVGRICVRLGW
 VPLQPLPESLQLHLLEMYPMLENIQKYLWPRLCKLDQRTLYELHYQMITFGKVFC
 TKSKPNCNACPMRAECKHFASAFASARLALPGPEEKSLVTSGTPIAAETFHQTYIS
 SRPVVSQLEWSNTCHGMNNRQPPIEEPASPEPEHETEEMKECAIEDSFVDDPEE

IPTKLNFEFTQNLKSYM QANNIEIEDADMSKALVATPEVASIPTPKLNKNSRL
TEHQVYELPDSHPLLEGFNQREPDDPCPYLLSIWTPGETAQSTDAPKSVCNSQEN
GELCASNTCFSCNSIREAQAKVRGTLI PCRTAMRGSPPLNGTYFQVN EVFADH
DSSRNPIDVPRSWIWNLPRTTVYFGTISIPIFKGLTTEEIQHCFWRGFVCVRGFDRT
SRA PRPLYARLHF PASKITRNKK SAGSAPGRDDE

5 SRA PRPLYARLHFPASKITRNKKSAGSAPGRDDE

SEQ ID NO:18

>DMTRICE novel 723 amino acid NH2 terminus;

MQDPGQWLPSQSTADFLYFSSIPPSQDFTSIETQPTTSAVSESEKESANSPVHEGNTCLVERISNDALTEVGVSSAGP
TEC1DLNKTPKRPPKKKKHPRVKLUDKDKPSKTPSATIPSTEKEVPSGRKYVVRKTSQPCPQAECASASHERSELKS
VKVSLRDFFGEVLEQESTQSGSQVPAVCTGPKRQSISIPTQDSDQSOLACHVVSSTSIHTSASQMVNAHLFPDPNMPPNG
VLDDLNNSNTSQLEHAKFVDSPARLFGSRIRQTSQGSNSLLEIYAGMSDRNDVPNLINSSIQTHSMSTDFAQYLSSQSAC
ERTVQMANQMLHGRHMPEFNITPSCHTERAELKEHLMHVVHAKAAMVMQGCPHSYRLAQANQLPFLPNHIEQYVQMNELSEL
VTNDTLYTASPFSQTGAQNQHNIQDGSMHHALDPRRENSASSGSWISLGVNFTQONQNGWASAGADAASHPAYSFPEH
KMRPTA1LNNYNPGVVGHFESTSDLSNNENAVASAINSNVTLDLAQRLIAKERSRASQMSIFRSKNDVMRSEMV
HQHQRPHGPGSACRESIEVDFDQKMGTMELTQLQSLMPNHNQPREKEYIIPQTCGSQLQSLERHDWMVKGNHLAGELHKQVTSQ
VVIQNSFCVTPDVLRGRTSGEHLRTLIAPIHTASTCKDTKLALSCOLESRDIIRPPVNPIGFSSADVPRTDNHQKVSE
ETV

20 SEQ ID NO:19

>DMTRICE nucleotide sequence from PAC PO489G09;

10261 aaatattgtc taaaatggata taaaatgttga aaatgttactt gagggaaagt gttaggtgcac
10321 gtggggctcc acaaattttc ttcaactatgtg cacccttttagt tatatatattt ttggcgaaga
10381 ggacaaagtc gctccgtta attttgagta agggccggog ggatatttat ttgtgtaaat
10441 gagcttccaa agaaaaacat gatagtgcat atgtatcc ttcttttttttttttttttttttttttttt
10501 ttccataactg ttttacagaa ttcatgttgc gtttgttgcata ctgttgcac teatattttt
10561 gtatattttt actaaatgoa aqgtgttgtt gttcgtagt ttatattttt ttttagtgcatt
10621 ttccatgtt ctgtatgtat ttgtatgtga tccatccggg tgctttttt attatgttg
10681 tgctggggcc cccacataggaa ggccctgttg atccaatgtg gtgtccgtcc agttgtaaat
10741 goatagatcc atagacaacta ctgggttaaa cttagagatg gaatgtgtga aaggaaatat
10801 gcaggatttt ggacaaatggc tgccctcaatc tcgaccactt gocgatctat attttccgg
10861 tattccca caatccacatg tccatgttcc tcataagggc cagactaaat cccttcgcag
10921 tgatctcgca gggaaatggat ctgtcaatcc ttccgttccct ctaataatggta ttgggttttt
10981 tggacccgtt agcaactatgtt ctggggtaac tgaatgttgg gaaatgttgg ctggggcacaa
11041 tgaatgttatt gacttggaaa acagacccaggc acggaaacccc aagaagaaaa agcagacggc
11101 aaagggtgtt aaggacgata aaccatggaa gacactttaat ttgttactt caataccctt
11161 aacacaaaggtt gtagaaaaatc catctggaaa gagaaaaatattt ttccggcaaga agacatctcc
11221 aggccccaaatcc ctccggaaa agggcatgttccatctgttgc atgtatgttgc tgaatgtcgt
11281 taacaaaggat ttggatctttt gttggaaatgtt actgttggatgg agtacacaaat ctggatctcc
11341 agttccgggtt gcacaaaatgtt gtacttggcc caagggttcaaa tcaatccctt etaccatccat
11401 aagatgttgc caaaaggatgtt gttggatggca ggggtttttt agacaaacgtt caatttcac
11461 ttccatgttgc catgttggta atccatgttgc ttccatctgttgcataatgttgc aaaaatggat
11521 atttgcgttgc ctccaaatattt ctactatgttgc ttccatggaaa aacatgttgc aattttgtgg
11581 cagttccggca gttttttttt gttccggaaa aagacacatca tcaggtaaaa atttttgtgg
11641 aaaaatgttgc ttccatgttgc catgttggaaa ttgttccatgttgc tcaatccatgttgc
11701 gacgttccatgttgc attttgcataatgttgc ttccatgttgc aaaaatggat
11761 aaaaatggaaaactt ccataatggccaa atccatgttgc ttatgttgc aaaaatggat

11821 tacacctagt cattgttattg aaagggtctgc attgaaggaa catttgaatc atgttcotca
 11881 cgaaaagcc gcagtatgc atggccaaat gcccataatg tacaggatgg cgaaaatcc
 11941 catccatccat ccaaatccata ttgaaggta tcaagtatgc gaaaatttg gtgaatttg
 12001 caccatcaaact gactatcaa ctgcgtatcc tttcgtatca actggagctg caaataggc
 5 12061 gataataat ggtatgttca tcgtatatac tgcatgtatc cctagaagag agatgtatgc
 12121 ttcaggatgt ctctttgtatc catttaggtt gaactttaac caacaaataat atggatggc
 12181 atctcgaggc gctggcgatg ctggagatcc acatgccccca tattttttag aacctcacaa
 12241 aagaATGagg acagcttcat ttaaaccat tccaaatggg gtctggggac attttttac
 12301 ctatctcaag gattgttcaaa ataatggg taaaaatgtg gcttcagcaaa tcaactcaaa
 10 12361 ctgttttaccc ctgtgtatgc cacaaggat gatagccgt gagaatcag gagttccca
 12421 aagaatgatc agttttagat catctaaaaa tgatatggg aacagatcag aatggtcca
 12481 tcaacatggc agacatgtcc cgcgtatgc tgcatgcagg gagttatgc aagtactgt
 12541 caaacatgtcc gggtcatga cagaagaact cacaaatata cttagtatgc caaataaccc
 12601 acaaaaggaa aataatattc cgcaactgtt aagtggccat cttagtctt tgaaatcatg
 15 12661 catgtttaaa gggataact tggcaggatgc attgtatcaag caagtaactt caccctcaatg
 12721 tggttattcag acaatccatc tggttacccc tcttgtatgtc ctccggagaa gaaacatgg
 12781 ggacatgttcc aacccatc tagtccaaac acatgtatgc aatgtatgg aactctgt
 12841 agctttaatgt tgtaactatgg agatgtttag agacattttt aggccctgtg taatccat
 12901 agggccatcc tctggcgatg ttccaaagac tgataaccat caagtcagg tttctgaga
 20 12961 aaccgttaccc gccaacttcc ctggagaacg aaaaatggg cgtcccgaaa aagagttaaa
 13021 acctgtgtgg aaaaaaccaactt ccatggccg tccaaaggaa gggaaaatgg ttgtggaga
 13081 atgttgcattt aaggatgttc acatcaatcc attgtttatc aagatgtactt catgttctt
 13141 tggttccatc ggacggggggg cttctgtgg aagagcaggta aagaaataa gagttggaga
 13201 aaacattttt cggatctatgg tatccctactt ggtatcttta gatatttttta ttccaaatgt
 25 13261 aaaaatgggttcc gacataaaactt aatccggaaa cctgtgtaca gatgttgcattt atgggtgtct
 13321 tggttccatc aatggatgtt tggttccatc tggttccatc gggggaaaatgg taaaagaaaa
 13381 aggttccgtt gccaatgtt atgttgcacc tgtaactgtt ttaatgttca gagatgttca
 13441 gggaccatgt atggatgtt gttgttgcggg aatggatgtt gatgttgcggg aatggatgt
 13501 tggaaagaaa aaaaatcc aagggtgtt tgatgttccattt attgtgttgcggg tgatgttgc
 30 13561 tcaaggatgtt totatccatc taaaatgtt tggttccatc aacatgtatgc gtttccatc
 13621 tggtgtatgtt gcccacatc tagtgcatttggg tccaaatccatc tcaatatttta ctgtgtaccc
 13681 aataatccatc tttccatccc aatatattca tggttgcattt cgtactgtcc cattagactt
 13741 gatgttgcgtt gttgttgcggg aacatccatc ttgttgcattt gatgttgcattt tggttgcggg
 13801 tggcggtttt tttccatc aaggatgtt gttgttgcattt gatgttgcattt tggttgcggg
 35 13861 acagaaatgtt tgccgaccatc ttcccgatgtt aataatgtcc agatgttccatc tggaaatgt
 13921 gactgtatcc gatgttgcggg gttatgttgcggg ttgttgcattt gatgttgcattt tggttgcggg
 13981 gtttccatc gatgttgcattt gttgttgcattt gatgttgcattt gatgttgcattt tggttgcggg
 14041 cggcaatgtt gatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt tggttgcggg
 14101 attgttgcattt gatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt tggttgcggg
 40 14161 ttatccatc aacatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt tggttgcggg
 14221 ggtatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt gatgttgcattt tggttgcggg
 14281 gactgtatcc gatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt tggttgcggg
 14341 ggttccatc gatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt tggttgcggg
 14401 attgttgcattt gatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt tggttgcggg
 45 14461 aacacatgtt gatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt tggttgcggg
 14521 gaatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt gatgttgcattt tggttgcggg
 14581 ttccatc gatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt gatgttgcattt tggttgcggg
 14641 aacatgttgcattt tggttgcattt gatgttgcattt gatgttgcattt tggttgcattt tggttgcggg
 14701 acatgttgcattt gatgttgcattt gatgttgcattt tggttgcattt gatgttgcattt tggttgcggg
 50 14761 attgttgcattt gatgttgcattt gatgttgcattt tggttgcattt gatgttgcattt tggttgcggg
 14821 atgaaacaca aagggtgtccatc tttccatc aacatgttgcattt tggttgcattt gatgttgcattt tggttgcggg

14881 agactacttc cctgtctgag cttcacacaa gaatcgataa ctgttttcaa ccattaagg
 14941 cagagaaaaat gtcatgttgc agggaaacagt cctttctgtaa aaattatctt tcaaggatg
 15001 aagctgaago tgcatttgtt aaacagcgtg gaacatcaaa ttgtcaagggt gataatactg
 15061 tcaggacaga gcaaaatggt ggtggaaaatt ctaatcagg atacagccaa caggatgata
 5 15121 atgttggtt tcaacaggcg acaaccaggta atcttttttccatcaacttca tgccaaacc
 15181 agaaagcaaa ttcttgcgtt ctacacggag ttctttccaa cttgtatagag aattctaaag
 15241 atgacaaaaaa gacttccccca aaagttcccg tcgatggato aaaagcaaaag aggccaagg
 15301 ttggggctgg taaaagaaaa acatatgtt gggatatgtt gagaaagaaat gtcttttaca
 15361 gtcatgttca taaaagaaaaa ttcccgatgtt ctaaggactc aattgttgg gaaacaataa
 10 15421 gacaacgaga ggttgaaggaa atatctgtca caatttgcgtt gcgaggaaatg aataacatgc
 15481 tggcagaacg gataaaatgtt agtatggcat aaaacatgtt acatggaaat ttgacataac
 15541 tcttagtcatca ttgtcatgttgc tgcattttca tatagtttgc ttatattttgtt ggaaatccca
 15601 gttttggatc aaccataactt atctttgtca attcatgttgc caggacttcc taacccgatt
 15661 ggttggatgtt catggatgtt cttgtatgttgc ttgttgcgtt ttttttttttgc
 15 15721 gaatgttgcgtt aactaaattt attttgcgtt aacatttgcgtt atgcattttgg cccttggca
 15781 ttcttataattt tgcatttttgcgtt accttgcgtt tgcttgcgtt tgacaaatggt atgttgcgtt
 15841 catgggttgcgtt aatgttgcgtt aatgttgcgtt tgatgttgcgtt gatgttgcgtt taatttgcgtt
 15901 ttttttttttgcgtt ttggatgttgcgtt taatgttgcgtt cttgttgcgtt cttgttgcgtt gcaaaatcc
 15961 aacgttggaaat ttgttgcgtt ctttttttttgcgtt ctttttttttgcgtt ttttttttttgcgtt
 16021 gggccatcttgcgtt ttttttttttgcgtt aatataattt aatataattt ttttttttttgcgtt
 16081 aaaaatgttgcgtt aatatttttatttgcgtt accatgttgcgtt gaaaccatgttgcgtt
 16141 ttgttgcgtt gtttttttttgcgtt aatataatttgcgtt ttttttttttgcgtt
 16201 gogatgttgcgtt attttttttgcgtt gggggacacat ttgttgcgtt gtttttttttgcgtt
 16261 cacacaggca ttgttgcgtt aatccatgttgcgtt ctttttttttgcgtt
 16321 agcccgatgtt ttttttttttgcgtt ttttttttttgcgtt
 16381 caatgttgcgtt taaaatgttgcgtt aatataatttgcgtt ttttttttttgcgtt
 16441 ttgttgcgtt ttttttttttgcgtt aatataatttgcgtt ttttttttttgcgtt
 16501 ttgttgcgtt ttttttttttgcgtt ccaggacttgcgtt ttttttttttgcgtt
 16561 aagtgttgcgtt ttttttttttgcgtt
 30 16621 ttgttgcgtt ttttttttttgcgtt
 16681 ttgttgcgtt ttttttttttgcgtt
 16741 ggggttgcgtt ttttttttttgcgtt
 16801 ctgttgcgtt ttttttttttgcgtt
 16861 ctgttgcgtt ttttttttttgcgtt
 16921 ttgttgcgtt ttttttttttgcgtt
 16981 ttgttgcgtt ttttttttttgcgtt
 17041 aatgttgcgtt ttttttttttgcgtt
 17101 ttttttttttgcgtt ttttttttttgcgtt
 17161 agcccgatgtt ttttttttttgcgtt
 40 17221 agcccgatgtt ttttttttttgcgtt
 17281 ttgttgcgtt ttttttttttgcgtt
 17341 ctgttgcgtt ttttttttttgcgtt
 17401 ttttttttttgcgtt ttttttttttgcgtt
 17461 caatgttgcgtt ttttttttttgcgtt
 17521 gacatgttgcgtt ttttttttttgcgtt
 17581 ttgttgcgtt ttttttttttgcgtt
 17641 ttgttgcgtt ttttttttttgcgtt
 17701 ttgttgcgtt ttttttttttgcgtt
 17761 aacatgttgcgtt ttttttttttgcgtt
 50 17821 ttttttttttgcgtt ttttttttttgcgtt
 17881 aatgttgcgtt ttttttttttgcgtt

卷之三

21001 ttgaaatttt cattgtcttc tgaatgtate ttacttggc atccataata ttccatatttg
21061 tcacgatata gfggtatagg acaaataaa catgggaatg attttttttt atccaaatc
21121 tactcatata attttcaataa gacaatataa aataatattt tgtagtgtcc
21181 tgtataaaa aatttgtcaca ccctatcgca cagcgcacaa aattttgtgc taccatggc
5 21241 agttgtactt ctatggatgat cttcatatgg aatcgaaataa aatgcoaatg tgaatgaaca
21301 aggatataatcata gtataccctc aaaatctctt gagaggatgt cgtaatataatg taggtgtataa
21361 taaaacaatgg tttccatatgg agggttttttt taaaaggatgt acaaacttgc tcaatataatgg
21421 taaaagtatgtt tttatccataa ggccatgttg cgacaaatgtt gttttttgtt gatgtactat
21481 ccctccgtcc cacaataataa gagatgttttgg gttttttgtt goaaatgttgg accacttgc
10 21541 tttttttttttt tttttttttt tttttttttt gacttactttt attttccaca
21601 gtactttaag tacaatctttt cgtttttttt atttgooaaa aaaaattgtat aagacgatgt
21661 gtccaaatgtt gtacggccaaa actccaaaatc ctttatattt tggggacggag ggatgtactta
21721 tggatgtccctt tttgtccaaatc gatgtcgtt aatctttttt tcaggatgtt ggtttttttt
21781 ttctttttttt cctaaatctttt tcaggatgtt tgggtgttgg aggttttttgg aggacatcaa
15 21841 gageacccatgg accactgtat gcaagactcc actttccagc aagaaaaattt accaggatca
21901 aaaaatctgc aggtttgttgc ccaggaaagag atgtatgttgc ggcatactgg aaaaaccatgg
21961 agggaaataaa gaggaggatc atatgtatcc cccagaatgttgc actgcacttgc aatggatctgc
22021 tgccaaatata gttgtcgttgc gaaatcttcaat tatttttttttgc catatcataatc tggatgttgg
22081 ttatgaaatccc agccactgtac gaaatctgttgc tgggtgttattt ctoggacac atgtatccatc
20 22141 atacagatcc tttttttttt gtttgcacgg aattttttatg cgggtgcacgt tggatgttgc
22201 agggagaaaaa ggcatgttgcg ggggttttttttgc ggttttttttgc ttttttttttgc
22261 cccatgttttttcatgttgc aaggatgttgc aatggatcttgc ttttttttttgc ttttttttttgc
22321 ttatgtatgttgc ttttttttttcatgttgc aatggatcttgc ttttttttttgc ttttttttttgc

25 SEQ ID NO:20

CORN(ZEA MAYS)DMT.1

```
>Corn DMT.1 660990 (688512 selclone ID);
EPDDPCYLLSIWTPGETAQSIDAPKTFCDSGETGRLCGSSTCFSCNNIREMQAQKVRG
30 LLIPCRTAMRGSPFLNGTYFQVNEVFADHCSSQNPIDVPRSWIWLPLRTVYFGTSVPTI
FRGLTTEEIORCFWRGFVCVRGFDRTVRAPRPLYARLHFPVSKVVRGKKPGAAAREEE
```

SEQ ID NO:21

```
>Corn DMT.1 cDNA 660990 (668512 selcone ID);
35 gaaccagatgatcccttgtccatatctttccatatggaccccaggtaactgcacaa
          tcgatcgatcccccaagac
          attctgtatcaggggagacggtagactatgtggaaagttaacatgcttagttcaa
          caatacgcagaatgcagg
          ctcagaagtcagaggaacactttgataccatgccaaacagcaatgagaggaagttcc
40 cacttaatgggacgtatTTT
          caagttaatgaggatTTGCTgaccattgtcaagtcaaaatccaattgtgtcccacga
          agttggattggacccccc
```

aagacgaactgttactttggAACCTcagttctacaatattcagaggTTAACGACTGA
 agagatacacacgatgttt
 ggaggaggattttgcgtggggcTTTgataggacagtgcggcaccaaggcccTTT
 atgcaaggTTgcattttcT
 5 gtcAGCAAGGTTGTTAGAGGCAAAAGCCTGGAGCAGCAAGAGCAGAAGATAATAGAAC
 attGAAGAAAATAGGGTG
 ctaaccAGATGAGGATGGATAGCCCCAAATGAGATGCTGACCCAATAGGTGCGCAAATCA
 CCTCCAAATTCTAACCAAT
 gacttccATCTGTAATGGCAATACCTGAAAACCTGTGATGGAGATGTTTGTGGC
 10 gACATGATCTCTAAATTAG
 attcgTCTTGGTAACAGCCTAGCTGTTCTGTTGAGTCGCATATTCTTATTCTGAAG
 ATCAATATAGCAAATGGG

SEQ ID NO:22

15 CORN(ZEA MAYS)DMT.2

>Corn DMT.2 371537 (441428 selclone ID);

 MITFGKVFCRKQPNANCAPMRSECKHFASAFASARLALPAPQEESLVKLSNPFAFQNSS
 MHAMNSTHLPRLGESIHSREFLPKNSEPIIEEPASPREERPPXTMENDIEDFYEDGEIPT
 IKLNMEAFQAQNLENCIKESNNELQSDDIAKVALVAIXTEXASIPXPK

20

SEQ ID NO:23

>Corn DMT.2 cDNA 371537 (441428 selclone ID)

 tATCAGATGATTACATTGGAAAGGTCTTTGTTACAAAAGACAGCCTAAATTGCAATGCA
 TGGCCAAATGAGGAGTGTAGTG
 25 caAGCATTGCAAGTCATTGCAAGTGCAGGCTTGCACTTCCTGCTCCCCAGGAGGA
 aAGCTTAGTGAAGTTGAGCA
 ATCCATTGCTTCAGAAATAGCAGCATGCTATGAATTGACTCACCTACCTGCC
 TTGAGGGAGGTATCCATTCA
 AGGGAGTTCTTCAGAAACTCAGAGCCAATAATCGAGGAGCCTGCAAGTCAGAGAG
 30 gaaAGACCTCCAKAAACCAT
 GGAAAGATGATATTGAAGATTTTATGAAGATGGTGAATCCAAACAATAAGCTTAACAT
 GGAAGCTTTGCAACAAACT

tggagaattgcattaaagaagaacaataacgactccagtctgtatattgcaaaagcat
tggttgcattarcactgaa
rcaqcttsattcctgkaccgaaaact

5 SEQ ID NO:24

Corn (*Zea mays*) DMT.3

>Caro DMT-3 218853;

MPRPKPKRKAPASPARHDPSPPEPYPSHASP CSAQC L VVRD ALLA F HGF PEEFA AFRV LRLG
GLSPN RDPRPSSPTVL D GLVTT L S QNTT DAIS RRAF ASLKA A FP S WQV V DEEG K RL ED
AIR CGGLAATKAARIRSMLRDVERRGKICLEYLREL SVDEVKKELSRFKGIGPKTVACV
LMFYLQKDDFPVDTHVLRITKAMGW VPATASREKAYIHLNNKIPDDLKF DLNCLFVT HGK
LCOSCTKKVGSDKRKSSNSACPLAGYCCIGEKLQLQ

SEQ ID NO:25

15 WHEAT DMT.1

>Wheat DMT.1 614028 (887053 selclone ID);

MRAECKHFASAFASARLALPGPEEKSLVTSGNPIASGSCQQPYISSLRNLQDWNANAH
HILDNRQPIEEEPASPEPEPETAEMRESAIEDIFLDDPEEIPTKLNFEFAQNKLKNYMQ
VNNITFEMEDADMSSALVAITTPFAASIPTPRLKNVSRLLTEHQVYELPDSHPLLEGYDQREP

20 DDP

SEQ ID NO:26

>Wheat DMT.1 614028 (887053 selclone ID);

tgcccaatgagagctgaatgcaggacttgcatttgcaggactgttagacttgtttctggaccctg

SEQ ID NO:27

Wheat DMT.2

>Wheat DMT.2 568842 (908118 selclone ID);

卷之三

NRVDESTVGAGDAASPKTRTRRKNTENFDWDKFRRQACADGHMKERKSERRDSVDWE
AVRCADVQRISQAIRGMMNVLSERIQEFLNRLVRDHGSIDLEWLDIPPDASKDYLSS
IRGLGLKSVECVRLLTLHHLAFTPVD

5 SEQ ID NO:28

SEQ ID NO:29

15 WHEAT DMT.3

>Wheat DMT.3 611792 (838515 selclone ID);
NRKVQNEVFDHKSSYDPIYVAREQLWKLERRMVYFGTSVPISKGLTTEEIQQCFWKGF
VCVGERFETGAPRPLCOHLHVAASKVPRSRNAAAAGI_NSDSAKASAP

20 SEQ ID NO:30

>Wheat DMT.3 611792 (838515 selclone ID);
aatcggaaaacaacttataatgggatattgcagaccacaaatctgactcgatcccatatacggttgcgggaa
gcaggatgttgcggaaaggatgttgcgggatctggacttttcgtgcggccatattcaaaaggcttaacactgaagaata
cgagtgttgcggaaaggatgttgcgggatctggacttttcgtgcggccatattcaaaaggcttaacactgaagaata
cggtcgccgtacaaagtgcgcggatcacgcacgcggcggcggcggatctggctgttgcggatctggatcc
tgatgtatcatcacacccgtatcgaccctgtgcgtatgggtacgcgtatgttgcgttcgcgggcwacaggcgttgc
aaccsctgcgcaraagaattatcatccgcgttgcgttgcgtatctgttgcgttagkacccctttTaaaatccctgtgc
tacccgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
traaaaactgttgcakacgcaccttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc

30

SEQ ID NO:31

Wheat DMT.4

>Wheat DMT.4 615131 (861906 selclone ID);

MRSECRHFASAFASARLALPAPQEKSLSVMSSNQFSQSGGMPTPYSTVLPQLEGSAQGQD
 FCTNNSEPIIEEPASPAREECPETLENDIEDYDPDTGEIPLIKLNLQAFAQNLENCIKES
 NMMDLGSDIAKALVAVSTGSASIPV

5 SEQ ID NO:32

>Wheat DMT.4 615131 (861906 selclone ID);
 tactttggaaagggtttctgtacaaaaaaaacaaggccaaattgcaatgctgtccaatgag
 aagcgaatgcaggcattcgcagaatgccttcgcagaatgcacggcttgcacttcctgcacc
 tcaggagaaaagtggatgtcgagcaatcaattcagttccagagtggatggcatgcc
 10 aactccatactaactgtgcttcctcagttgaggaaagtggccaggcacaggattttg
 cactaacaattcagagccaattattgaggagccagcaatgcacggcaatgtcc
 agaaaactcttggaaaatgata
 ttgaagattacgatccagatactggtaatcccactaattaagcttaacttgcgaagctt
 ttgctcagaacttggaaaactgcattaaagaaaagcaatatggatcttgggtctgtatgata
 15 tcgcgaaagcacttgtgttagactggatecagctcaattctgtccc

SEQ ID NO:33

Soybean(Glycine max)DMT.1

>Soy DMT.1 449122 (557119 selclone ID);
 20 MDSLDWDAVRCADVSEIAETIKERGMNNRLADRIKNFLNRLVVEEHGSIDLEWLRDVPPDK
 AKEYLLSIRGLGLKSVECVRLLTLHHLAPPVDTNVGRIVAVRLGWVPLQPLPESLQLHLLE
 LYPVLESIQKYLWPRLKLDQETLYELHYQMITFGKXFCRKSKPNCACPMRXECRHFAS
 AFASARFALPGPEQKSIVSTTGNSVINQNPSEIIQLHLPBPENTAQEDEIQLTEVSQL
 ESKFEINICQPIIEEPRTPPECLQESQTDIEDAFYEDSSEIPTINLNIEEFTNLNQ
 25

SEQ ID NO:34

>Soy DMT.1 449122 (557119 selclone ID);
 aaaaaatthaakagacaaggaaacaagaaaaaggatgaytttgactggatagttaaagaattg
 aagcacaggactaaggctggaaaagagaaaaacagatacacccatggattttggactggatctgtcagat
 30 gtcagtgaaatcgctgagacccatcaaagaaaggccatgcacaacacaggctgcagatcgatataagaatttcataatcgatgg
 tgaagaacatggaaagcatgtggatctgtggatggactgttccacactgcacaagaaaagaataacttgcgcataaggg
 ttggactaaaaatgtggatgtgtggctttaactgcacatcttgcctccggtagacacaaaatgtcgccagctatgc
 gtacgactggatggatggccctcacagccactgcctgactgcagttgcacatcttgcataatgttaccagttggatgtcaata

caaaaatatctctggccctcgactatgcacgttagatcaggaaacactatacgatgactacattaccagatgattacatttggaaaggkc
ttctgtacaaaaggcaaccataatgtatgcattgcataatggaggagaat

SEQ ID NO:35

5 SOYBEAN(*GLYCINE MAX*)DMT.2

>Soy DMT.2 387990 (473695 selclone ID);

MRMTIDLVSQQSLTARLQLSILKDKLKIQCRKARGLDGFRNESSKIDSSPVKLRSREHKG
EKKNNFDWDSLRIQAEAKAGKREKTENTMDSLDWDAVRRADVSEIANAIKERGMNNMLAE
RIQSFLNLLVDKHGGIDLEWLDRVPPDQAKEFLLSIRGLGLKSVECVRLLTLLHHLAFPVD

10 TNVGRIAVRLGWVPLQPLPESLQLHLLLELYPVLESIQKYLWPRLCKLDQRTLYELHYQLI
TFGKVFCFKSK

SEQ ID NO:36

> Soy DMT.2 387990 (473695 selclone ID);

15 gaaaagataggatcatctcgatagcaactcagaaatagaagacacctgtctagcgctgcc
aagtacaatagttattataa

tagaattcttcagtcgttagcttttagaaatggcaagttcaaccatgttgcattgaagttaa
cagtcaaaagaagcaatcaa

ctgagaacttaggagatacatgtgatcagtctatagacatgaagcatgacaacctggcag
20 aaaacttggaaaaatcgat

gttactaaggctccgcagaaggcacccatccaatggatatactttaaaataacccca
aactcaggagtacttgaggat

taactgttatgatcctctaaaaatagaagtcccatacgatggctctcaagggtaaaga
tgagaatgacaatagatcta

25 gttcccaacagactgtcgactgccaggctgcattgtccattctcaaggacaaactgaag
atccaatgcagGaaagcaag
gggactagatttggtaggaatgaaagcagtaagatagattttccctgtaaaattaag
gagcaggaggcatggaaaag

30 agaaaaagaataacttgattggatagttaaagaatacaagcagaagctaaaggcaggga
aaagagaaaagacagagaac
accatggactccttgactggatgttagacgcgcagatgtcagtggaaattgcaat
gcaatcaaagaaaaggggcat

SEQ ID NO:37

SOYBEAN (*GLYCINE MAX*) DMT.3

```
15 >Soy DMT.3 657152 (546665 selclone ID);
INQAEQQTEVIRQLEAKSEINISQPIIEPATPEPECSQVSENDIEDTFNEESENPI
KLDIEEFTLNLQNYMQENMELQEGERMSKALVALHPGAACTPTPLKNVSRRLTREHYVYEL
PDSHPLLNGWNKREPDDPGKYLLAWPGETABSIQPPESKCSSQEECGXLNCNENECAFSC
NSFREAXFXDSXRDTPDTSNSXXXGAFH
```

20 SEQ ID NO:38
>Soy DMT.3 657152 (546665 selclone ID);
tataaaaccaaggcagaacctcaacaaacagaagtgtatcaggcaactagaagcaaaatctga
aatcaacatcagccaaccta
25 ttatttgaagagccagcaactccagagccagaatgtcccaagttatccgaaaatgtatag
aggataccttcaatggagaa
tcatgtgaaattcccaccatcaaactagacatagaagagttcacatggaaacttacaaaac
tatatgcacaaaaatcatggaa
acttcacaaagggtgaaatgtcaaaaggccttggtgcctacatccaggtgcacatgcac
30 tcc tacacccaagctgaaga
atgtgagccgggttgcacacagagcattatgtttatgaactccctgattcacatccccctc
tgaatgggttgcacaaaggcga
gaacctgtatgcacccaggcaatacctttagctatatggactccaggggagacagcagat
tctatacagccaccaggaaag

caaatgcagctctcaggaatgtggccggctctgtaatgagaatgaaatgttttcatgaa
 cagtttccgtgaagcaagg
 tcacagatagttcgagggacactcctgataccatgtcgaacagctwtgaragggag

5 SEQ ID NO:39

SOYBEAN(*GLYCINE MAX*)DMT.4

>Soy DMT.4 432980 (663678 selclone ID);

EAASIPMPKLNVSRLRTEHCVYELPDTHPLLQGWDTREPDDPGKYLLAIWTPGETANSI
 QPPESKCSSQEECGQLCNENEFCFSNCNSFREANSQIVRGTLV

10

SEQ ID NO:40

>Soybean DMT.4 432980 (663678 selclone ID);

agaagctgcttccattcttatgcccagctaaagaatgtgagccgattacgaacagagca
 ttgtgttatgaactccag

15 atacgcatcctttctccaagggtggcacacacgagagcctgatgatccaggcaaatac
 ttcttgctatatggactcca

ggtagacagcaaattctatacagccaccagaaagcaaattgcagctctcaagaagaatgt
 ggccaaactctgtaatgagaa

20 tgaatgtttctcgtaacagttccgtgaagcaaattctcagatagtttagagggacact
 cctggctctgaatgcattatca

aaatcattgttttaaccatatgttagctactaattttatacattatggAACAGGGAG
 ggaatacatctccatagaaa

ttcaaggcattataatagactgacttgaatttatgataaatatgagcagataccatgt

SEQ ID NO:41

>**Medicago 6654943;**MELQEGEMSKALVALNQEASYIPTTKLNVSRLRTEHSVYELPDSHPLLEGWEKREPDDP
GKYLLAIWTPGETANSIQPPDRRCQAQDCGQLCNEEECFSCNSFREANSQIVRGTLI5 RTAMRGSFPLNGTYFQVNNEVFADEHSSLNPISVPRSLIWNLDRTVHFITSVTSIFKGLA
TPEIQQCFWRGFVCVRSFERSTRAPRPLMARLHFPA

SEQ ID NO:42

>**Medicago 6654943 EST306265**10 GAGAACATGGAACCTCAAGAAGGTGAAATGTCAAAGGCCTGGTTGCTCTAAATCAAGAA
GCTTCTTACATTCCCTACAACGAAGCTGAAGAACGTCAGTGAGTCGGTTGCGCACAGAGCATTCT
GTTTATGAACTCCCAGATTCTCATCCTCTTCTGGAAGGGTGGAAAAGCAGAGAACCTGAT
GATCCAGGAAAATACCTTCTAGCTATATGGACGCCAGGTGAGACTGCAAATTCTATAACAG
CCACCAAGACAGAACGATGCAGCGCTCAAGATTGTGCCAACCTGTAATGAGGAGGAATGT
15 TTTTCGTGCAACAGCTCCGTGAAGCAATTACAGATAAGTTGAGGGACAATCTGATA
CCATGTCGAACAGCTATGAGAGGGAGCTTCCGCTAAACGGAACCTATTTCAAGTCAAT
GAGGTTTTGCAGACCATGAATCAAGCTTAATCCGATTAGCGTTCCAGAAGTTGATA
TGGAACCTTGATAGGAGGACAGTCATTTGGAACCTCCGTAACAAGCATATTCAAAGGT
TTAGCAACACCAGAAATTCAACAGTGCTCTGGAGAGGGTTGTCTGTGCGGAGCTT
20 GAAAGGTCAACGAGAGCACCCCGTCCTTAATGCCAGACTGCATTCCCAGCAAGC

SEQ ID NO:43

>**Tomato 12624037;**

MELQEGEMSKALVALNQEASYIPTTKLNVSRLRTEHSVYELPDSHPLLEGWEKREPDDP

25 GKYLLAIWTPGETANSIQPPDRRCQAQDCGQLCNEEECFSCNSFREANSQIVRGTLI
RTAMRGSFPLNGTYFQVNNEVFADEHSSLNPISVPRSLIWNLDRTVHFITSVTSIFKGLA
TPEIQQCFWRGFVCVRSFERSTRAPRPLMA

SEQ ID NO:44

30 >**Tomato 12624037 EST469495**GCTTGAGAAAGGAAGTCCAATCAAAGAGTGGAAAAAGAAAAGAAGCAAGGATGCAATGG
ACTCATGAACTACGAAGCAGTCAGAAGTGCAGCAGTAAAGAAATTCTGATGCTATTAA
AGGAACCGAGGGATGAACAAACATGCTGGCAGAGCGAATTAGGACTTCCTCGATAGACTGG

TGAGGGATCATGGAAGTATTGACCTAGAACATGGTTGAGAGATGTGGCCCCAGACAAAGCGA
 AAGAGTATCTTTGAGTATTCGTGGACTGGGCTGAAAGTGTAGAACATGTGTGCGGCTAT
 TAACACTTCAACCTTCTTCCAGTTGACACAAATGTTGGACGAATAGCTGTGAGAT
 TAGGATGGGTCCTCCAACCACCTCCTGAGTCCCTGCAGTTGCATCTTCTTGAACGTG
 5 ATCCAATTCTGGAGTCATTAGAACAGTATCTCTGCCACGACTCTGCAAGCTCGATCAGA
 GAACACTGTATGAGTGCACCTACCATGATTACCTTGAAAGGTTTCTGCACCAAAA
 GTAAGCCTAACTGTAATGCATGCCACTGAGAGCTGAATGCAGACACTTGCTAGTGCCT
 ACGCAAGTGAAGACTGCCCCCTTGCCCCAGAGGAGAAAGAGTATAGTGAGTTAGCAG
 TTCCGATCCCTAGTGAGGGAAATGCAGCTGCCGATTCAAGCCCAGCTATTACCCCCAG
 10 AGCTGAAGTAGGGATGGCGTACCCATATGCTCCAATTG

SEQ ID NO:45

>Barley 13256964;

MASETETFAFQAEINQLLSLIINTFSNKEIFLRELISNASDALDKIRFESLTDKSKLDA
 15 QPELFIFIIPDVKATNTLTLIDSGIMTKSDLVNNLGTIARSGTKDFMEALAAGADVSMIG
 QFGVGFYSAYPCAERVXVTSKHNDDEQYGGEXQAGWLLYCGHVILLESFGGVLRSPSTS
 RTNSWSTLERRAFKDGLGKNTPSS

SEQ ID NO:46

20 **>Barley 13256964 HVSMEI0014B12F**
 CGAGAACCCCGCTCAAAGCCCTAACCTAGGCCATCCCTCTCCCTCCCTCAACCCCTC
 GTCGACTCCCGCGCGCCCTGCGTTCAGGAGCTCCGCTGCCGGCGGCCATGGCCTCA
 GAGACCGAGACCTTCGCTTCAGGGGGAGATCAACAGCTCTCGCTCATCATCAAC
 ACCTTCTACTCCAACAAGGAGATCTCCCGCGAGCTCATCTCAACGCCCTCGATGCG
 25 TTGGATAAGATCAGTTGAGAGCCTCACTGACAAGAGCAAGCTGGATGCTAGCCAGAG
 CTGTTCATCCACATTATCCCTGACAAGGCCACCAACACACTCACCCCTATCGACAGTGGC
 ATTGGTATGACCAAGTCAGACCTCGTAACAACCTGGTACCATGCAAGGTCCTGGCACC
 AAGGATTTCATGGAGGAGATTGGCTGGTGGCGATGTTCCATGATTGGTCAGTTGGT
 GTTGGTTCTACTCTGCTTACCCCTGTGCTGAGAGAGTCGNTGTGACCAGCAAGCACAAC
 30 GATGAGGAGCAGTATGGGGGGAGTNCCAGGCTGGTGGCTTACTGTGGACACGTG
 ATACTCTGGAGAGGCCCTTGGAGGGGACTAAAGATCCCCCTACCTCAAGGAGCAAC
 AGTTGGAGTACCTGGAGAGGCCCTTAAAGGATTGGGGAAAAACACTCCGAGTTCA
 TAACTTTTCATCTCCTCTGGACGGGGAAAACCCCTGAAAGAATTGGCTGGAAA

GTGGGTGAAAAATGGGTCCTGGGGGGCCCGTTGAGGGATTGGTCACATAAACAA
ACTATCGTCTTCTATCTTAGCACCTAATAGTCCTCACATGAG

SEQ ID NO:47

5 >**Corn 1BE511860;**
LLEGFEQREPDDPCPYLLSIWTPGETAQSIDAPKTFCDSGETGRLCGSSTCFSCNNIREM
QAQKVRGTLIIPCRTAMRGSFPLNQTYFQVNEVFADHCSSQNPIDVPRSWIWDLPRRTVY
FGTSVPTIFRGLTTEIQRCFWRGFVCVRGFDRTVRAPRALYAR

10 SEQ ID NO:48

>**Corn 1BE511860 946063H01.Y1 946 -**
TATGAAC TGCCAGATTCA CACGCCTTCTGGAA GGATT CGAACAGAGAGAAC CAGATGA
TCCCTGTCATATCTTCCATATGGACCCCAGGTGAAACTGCACAAATCGATCGATGC
CCCCAAGACATTCTGTGATT CAGGGAGACGGTAGACTAT GTGGAAGTCAACATGCTT
15 TAGTTGCAACAATATA CGAGAA ATGGCAGGCTCAGAA AGTCAGAGGAACACTTTGATACC
ATGCCAACAGCAATGAGAGGAAGCTTCCCAC TTAA TGGGACGTATTTCAGTTAATG
GGTATTGCTGACCATTGCTCAAGTCAAATCCAATTGATGTCCACGAAGTTGGATTTG
GGACCTCCAAAGCGAACTGTTACTTTGGAACCTCAGTTCCATAAATTCAGAGGTTT
AACGACTGAAGAGATAACA CGATGCTTTGGAGAGGATTGTTGCGTGAGGGCTTG
20 TAGGACAGTGCGGGCACCAAGGGCCTTTATGCAAGG

SEQ ID NO:49

>**Cotton 11206330;**
MQGNMELQEGDLSKALVALNPDAAS IPTPKLNVSRLRTEHYVYELPDKHPLLKQMEKRE
25 PDDPS PYLLAIWTPGETANSIQPPEQSCGSQEPGRLCNEKTCFACNSVREANTETVRGTT
LIPCRNAMRGFSFLNGT

SEQ ID NO:50

>**Cotton 11206330 GA_EB0023J04F**
30 CTCCGCCAGTGCTCATACTTGCTTAAAGTAGGGCTTAATGTTGGCAACAAATGAACCTATCA
TTGAGGAGCCTGCAACACCTGAACAGAGCATGCAGAAGGATCAGAGAGTGATATTGAAG
ATGCAAGCTATGATGATCCAGATGAAATTCCCACAATAAAACTCAACATTGAAGAGTTCA
CAGCAAACCTACAGCATTACATGCAGGGCAATATGGAACTCCAAGAAGGGACTTGTCAA
AGGCTTAGTAGCTTGAATCCTGATGCTGCTTCTATCCCTACTCCAAAATTGAAGAATG

- TAAGCAGGCTACGAACAGAGCACTATGTATATGAGCTCCAGATAAACATCCTCTTGAA
 AACAGATGGAAAAGCGGGAACCTGTGATGATCCTAGCCCCATCTTCTTGCATATGGACAC
 CAGGTGAAACTGCAAACCAATTCAACCACAGAACAAAGTTGGGTCCAAGAACAG
 GAAGACTGTGCAATGAGAAGACCTGCTTGCACAGTGTAAAGAGAACGACTAACACTG
 5 AGACAGTCGAGGAACCATCTTGATACCTGTAGAAATGCAATGAGAGGAAGCTTTCCCT
 TTAATGGGACTTAATTTCAGTTAATGAGGTCTTGCAGATCATGAATCAAGCCTCAA
 CCCAATCGACGTTCCAAGGGAAATGGATTGGAAATTAAACAAGAACGAACTGTATACTTG
 GAACATCCTGGTTCATCAATATTAAAGGACTTTCGACGAGGGAA
- 10 SEQ ID NO:51
>Soybean 5606759
 MGWVPLQPLPESLQLHLELYPVLESIQKYLWPRLCRKLDQETLYELHYQMITFGKVFC
 SKPNNCACPMRAECRHFASAFASARFALPGPEQKSIVSTGNNSVINQNPSEIISQLHLPP
 PENTAQEDEIQLTEVSRQLESKFEIYICQPIIEEPRTPEPECLQESXTDIEDAVYEDSS
- 15 SEQ ID NO:52
>Soybean 5606759 SB95C12.
 ACGAGCTTCCCGTAGACACAAATGCGACGTATTGCCGTACGACTGGATGGTGCCT
 CTGCAGCCACTGCCCTGAGTCACTGCAGTTGCATCTCTAGAATTGACCCGGTGTGGAG
 20 TCAATACAAAAATATCTCTGGCCTCGACTGTGCAAGCTAGATCAGGAAACACTATATGAG
 CTACATTACCATGATTACATTGGAAAGGTCTCTGTACAAAAAGCAAACCAAATTGT
 ATATGCAATGCCAATGAGAGCAGAAATGTAGACACTTGTAGTGCAATTGCAAGTGCAGG
 TTTGCAGTGCCTGGACCAGAGCAGAAAGAGTATAGTTGACAACTGGAAATAGTGTGATT
 AACCGAACCCATCTGAAATCATCAGTCAGTTGCAGTGCCTCCACCTGAGAACACAGCC
 25 CAAGAAAGATGAAATTCAACTAACAGAAGTGAGCAGACAATTGGAATCAAATTGAAATA
 TATATTGCCAACCTATCATTGAAGAGCCCAGAACCTCAGAGCCAGAACATGCTGCAAGAA
 TCACANACTGATATAGAGGATGCTGTATGAGGATTCAAGTG
- SEQ ID NO:53
 30 **>Wheat 12019155**
 MFHCHGTRGSDLGFDLNKTPEQKAPQRKRHPKVKEAKPKSTRKPATQKTQMKENPHKK
 RKYVRKTAATPQTNVTEESVDSIVATKKCSRRAINFLEHNKYASQSTISQCQEIDHRNE
 KAFNTTSDHKAKEPKNTDDNTLLLHEKQANNFQSE

SEQ ID NO:54

>Wheat 12019155

AACAGTCAGGACAAGGCAACAAGATCAGCAGTCAGGACAAGGGCAGCAACCAGGGACAAA

5 GGCAGGCCAGGGTACTACTCAACTTCTCGCAACAATTAGGACAAGGCCAACCAAGGTACT
 ACCCAACTTCTCGCAGCAGCCAGGACAAGAGCAGCAGCCAAGACAATTGCAACAACCAG
 AACAAAGGGCAACAAGGTCAAGCAGCCAGAACAAGGGCAGCAAGGTCAAGCAGCAAAGACAAG
 GGGAGCAAGGTCAAGCAGCCAGGACAAGGGCAGCAACAGGGCAGCAACCAGGGACAAGGGCAGC
 CAGGGTACTACCCAACTTCTCGCAGCAGTCAGGACAAGGGCAACCAGGGTACTACCCAA
 10 CTTCTCCACAGCAGTCAGGACAATTGCAACAACCAGCAGCAAGGGCAGCAACCAGGGACAAG
 AGCAACAAGGTCAACAGCCAGGACAAGGGCAGCAACCGGGACAAGGGCAAGGCCAGGGTAC
 TACCCAACTTCTCGCAGCAGTCAGGACAAGAGCAACAGCTAGAACAAATGGCAACAGTC
 GGACAGGGCAACCCAGGGCACTACCCAACTTCTCGTGCAGCAGGCCAGGACAAGGGCAAC
 AGGGTACTACCCAACTTCTCACAAAGATAAGGACAAGGGCAGCAGCCAAGAACAAATTG
 15 ACAACCAACACAAGGGCAACAANGGCAGCAACCAAGGACAANGGCACAAGGTCAACA
 GCCCANGAAAAAGGCAACAAAGGTCAAGCAACCAAGNACAAGGGCAGCAANCCAGGAC
 AAGGGCAGCCANGGTCTACCCAACTTTTGAGCAAGTCANGGAAAGGGGACCANC
 CNAGGANAAATGGNACCAACCCAGNACAAGGACAACCCGGGTCTCCCCAAANTTTTN
 CN

20

SEQ ID NO:55

>Tomato 8106032

MSLAHFPLKTDSTQKHEGNTGIIIEPEECATDPNVSVIRWYEDQPQNPQSTHCQDSSGVYN
 TDSNEEKPAVNDSESENSTECIKSAECSVILQSDSSREGSDLYHGSTVTSSQDRKELND

25 LPSSPSSVVSSEISAVIQASEGTDSSNFCSSTSFLKLLOMAGTSGAQGTRCTEHLHNQHK
 GNXGQQPRTXGNKVNSPKKKATKVKQPXTRGSXPQGQPKSYPTXFEQVXEKGHXRXNG
 XHPXQGQPRVFPKXF

SEQ ID NO:56

30 >Tomato 8106032 EST356474

CTCGTGCCTGGTTGGGGTATATCTTACACAGAACATGTTAGTCAGATCACCTTCTAGTTCTGCA
 TTCACTGTCAGTCGCTGCCACTTCTCTGAAAACAGACAGTACTCAGAACAG
 AATACAGGTATTATAATTGAAGAACCTGAAGAGTGTGCAACAGACCCCAATGTTCCATC
 AGATGGTATGAAGATCAACCAAATCAGTCACCCATTGTCAGGATTCTCAGGAGTCTAT

AATACAGATTCAAATGAAGAAAAACCAGCTGTCATGACTCTGAATCAAGTGAAAATAGC
 ACAGAATGCATAAAATCAGCAGAATGTTCTGTAATTCTGCAATCAGATTCTCTAGAGAA
 GGCTCAGATCTGTATCATGGATCAACAGTTACAAGATTCCCAAGATCAGAAAAGAGTTGAAT
 GATTTGCCCTCTCTCCGAGTTCTGTTCTCTGAGATCTGCTGTTATTCAAGCT
 5 TCAGAAGGAACTGACTCAAGCAACTTTGCAGCTCCACTTCTTTTGAAAGCTATTACAG
 ATGGCAGGAACCTCAGGAGCACAGGAACCAGGTGCACTGAACATCTAC

SEQ ID NO:57

>Corn **IAW042334**:

10 DAHPLLQQILGLDQREHDDPTPYLLAIWTPDGIKEITKTPKPCCDPQMGGDLCNNECHNC
 TAEKENQSRYVRGTLILVPCRATMRGSFPLNGTYFQVNNEVFADHRSSHNPISHVEREMLWNL
 QRMMVFFGTSVPTIFKGLRTEEIQQCFWRGFVCVRGFDMETRAPRPLCPHLHVTARPKA

SEQ ID NO:58

15 >Corn **IAW042334** 614027C01.y1 614 -

GAATTCGGCACCAGCAGATGCACATCCACTTTACAACAGCTAGGACTTGACCAACGGGA
 ACATGATGATCCTACCCCATACTTATTGGCCATATGGACACCGAGATGGAATAAAGGAAT
 AACTAAGACACCAAAACCATGCTGACCCCTAAATGGGAGGGGATTATGCAATAATGA
 AATGTGCCACAATTGACTGCGAGAGAAAACCAATCTAGATATGTCAGAGGCACAAT

20 TCTGGTTCTTGTCAAACAGCTATGAGGGTAGTTCCCACHTTAATGGCACTTACTTCA
 AGTCAATGAGGTATTGCTGACCAAGACATCTAGGCCAACACCAATCCATGTGAAAGGGA
 GATGCTATGAACTTGCAGGCGCATGGCTTTTCGGGACTTCAGTACCCACCATATT
 CAAAGGTCTAAGAACAGAAGAAATACAACAATGCTCTGGAGGGGATTGTCTGTGCG
 AGGATTCGACATGGAGACTAGAGCACCAAGGGCTCTGTGCCCAATTGACGTTATAGC

25 AAGGCCGAAAGCCCCCAAGACAGCAGCAACTGAGCAAGTACTCTAATCAGCAAAG

SEQ ID NO:59

>Corn **AW076298**

PCRTAMRGSFPLNGTYFQVNNEVFADHCSSQNPIDVPRSWIWLPLRRTVYFGTSVPTIFRG
 30 LSTEQIQFCFWKGFCVVRGFEQKTRAPRPLMARLHFPAASKLKNKLTTTEIQQCFWRGFV
 CVRGFDRTVRAPRPLYARLHFPAASKVVRKG

SEQ ID NO:60

>**Corn AW076298** 614065C03.Y1 614 -

CGGCCCCAGACCATGCCGGACAGCAATGAGAGGAAGCTTCCCACCTAATGGGACATATT
TCAAGTTAACGAGGTATTGCTGACCATGTTCAAGCCAAAATCCAATTGATGTCCCACG

5 AAGTTGGATATGGGACCTCCCAAGACGAACGTGTTACTTTGGAACCTCAGTCCTACAAT
ATTAGAGGTTAACGACTGAAGAGATAACAACAAATGCTTTGGAGAGGATTGTTGTGT
GAGGGCCTTGATAGGACAGTAAGGGCACCAAGGCCCTTATGCAAGGTTGCATTTC
TGCCAGCAAGGTTGTTAGAGGCAAAAAGCTGGAGCGCAAGCGTCGAAGATAATAGGT
ACATCGAAGAAATATAGAGGAGCTAACAAACGGATGGATAGCCCTAAATGAGATGCTGA
10 CCCAATAAGTCGCCAATCACCTCCAAGTCTAACCCAATTGAGGCGACATGACCTG
TTAAATTATGTTCCATCTATGGTAACAGCTTAGATGTTGTGAGTCGCATATTCTTA
CTCTGAAATTCAATATAGCAAATGAAAAAAACACAGTCGCATAGTCTAGTTCTAATTGTA
CCTGTGAGTGGAAATCAGTTGTTGACAACATGAAGATGGG

15 SEQ ID NO:61

>**Corn BE639158;**

KNSEPIIEEPASPREERPPETMENDIEDFYEDGEIPTIKLNMEAFAQNLENCIKESNNEL
QSDDIAKVALIASTEASIPVPLKNVLRLRTEHYVYELPDAHPLLQQLGLDQREHDDPT
PYLLAIWTPDGIKEITKTPK

20

SEQ ID NO:62

>**Corn BE639158** 946021E09.Y1 946 -

TGAGCTGCATTATCAGATGATTACATTGAAAGGTCTTGTACAAAAGACAGCCAAA
TTGCAATGCATGCTATGAATTGCACTCACCTACCTCGCCTGAGGGGAGTATCATTCAA

25 GGGAGTTCTCTTAAGAACATTGAGAGCCAATAATCGAGGAGCCTGCAAGTCCAAGAGAGG
AAAGACCTCCAGAAACCATGGAAAATGATATTGAAGATTTTATGAAGATGGTAAATCC
CAACATAAAAGCTAACATGGAAAGCTTGCACAAAACCTGGAGAATTGCAATTAAAGAAA
GCAATAACGAACTCCAGTCTGATGATATTGCAAAAGCATTGGTTGCTATTAGCACTGAAG
CAGCTCGATTCCCTGTACCGAAACTAAAGAACATGTGCTTAGGCTTCAACAGAACACTATG
30 TGTATGAGCTCCAGATGCACATCCACTTTACAACAGCTAGGACTTGACCAACGGGAAC
ATGATGATCTACCCCATACTTATTGGCATATGGACACCAGATGGAATAAGGAAATAA
CTAACGACACCAAAACCATGCT

SEQ ID NO:63

>**Corn T25243;**NHQPIIEPLSPECETENIEAHGAIEDFFCEESDEIPTINLNIEEFTQNLKDYMQANNV
EIXYADMSKALVAITPDAASIPTPKLKNVNRLRTEHQVYELPDSHPLLEGFEQXEPPDDC

5 PYLLSIWTPGELHNRSMP

SEQ ID NO:64

>**Corn T25243;**CTGGTAAATCATCGCCAATCATCGAGGAACCACTGAGCCCAGAATGTGAAACTGAAAATA
10 TAGAGGCACATGAGGGTGCATTGAGGATTCTTTGTGAGAAATCTGATGAAATTCCATA
CCATTAATCTTAAATCGAGGAGTTCACACAAAATTGAAGGACTATATGCAAGCAAACA
ATGTTGAGATTGANTATGCTGACATGTCAAAGGCATTGGTGCATCACGCCCTGATGCTG
CTTCATTCCAACCTCAAAGCTCAAGAATGTCAATCGTCTGAGGACAGAACACCAAGTTT
ATGAACTGCCAGATTCACACCCCTTCTGGAAGGATTGCAACAGNGNGAACAGATGATC
15 CCTGTCCATATCTCTTCCATATGGACCCAGGTGAACTGCAATCGATCGATGCCCA
AA

SEQ ID NO:65

>**Corn AW453174;**FQGNEVFADHCSRQNPIDGPRSWIWDLPRRTGYFGTSGPTIFRGLTSEEIQRCFWRGFVC
VRGFDRTVRAPRPLYARLHFVSKVVRGK

SEQ ID NO:66

>**Corn AW453174 660032D01.Y1 660 -;**CATGCCAACAGCAATGAGAGGAAGCTTCCCCTTAATGGACGATTTCAAGGTAATGAGT
GGTATTGCTGACCATGCTCAAGGCAAATCCATTGATGGCCACGAAGTTGGATTG
GGACCTTCCAAGACGAACTGGTACTTGGAACCTCAGGTCTACAATATTGAGGGTT
AACGACTGAAGAGATAAACGATGCTTGGAGGAGATTGTTGCGTGAGGGGCTTGA
TAGGACAGTGGGGCACCAAGGCCCTTATGCAAGGTTGCATTTCCTGTCAGCAAGGT
30 TGTTAGAGGCAAAAGCCTGGAGCAGCAAGAGCAGAAGAATAATAGAACATTGAGAAAT
ATAGGGGTGCTAACAGATGAGGATGGATAGCCGAAATGAGATGCTGACCCAAATAGGTC
GCCAAATCACCTCAAATTCTAACCCAAATGACTTCATCTGTAATGAATGGCAATACCTT
GAAAACCT

SEQ ID NO:67

>Corn BE509759;

NGTYFQVNEVFADHRSSHNPIHVEREMLWNLQRMMVFFGTSVPTIFKGLRTEEIQQCFCWR
GFVCVRGFDMETRAPRPLCPHLHIARPKARKT

5

SEQ ID NO:68

>Corn BE509759 946021E09.X1 946 -

TGGCATTCTAACATGGACTAACAGCTAGATGCTAATTACATACAGTAGATCTGAAACAAAAAGTAAAATTATGGTGC
TTCCGTGATCTTCATAGTCCTCGTCTCAGAAAACAAACAGTCTCGGACCCCATCCATGGCTTAATTCTCAAACAT

10

GGCTCTTTTTAGGCAGGAAGATAATGATTCCATGAGTCAGGTCTGAGAGCTATTGATGTCATATCACAAACATGATG
TTCAAAACTGATATTTGCTGATTAGAGTACTTGCTCAGTTGCTGCTGCTTGCGGCCCTCGGCCCTGCTATAATG
TGCAGATGGGGCACAGAGGCCCTGGTGCTCTAGTCATGCTCCATGCAATCTCGCACACAGACAAATCCCCTCAGAAGCA
TTGTTGTTATTTCTCTGTTCTAGACCTTGAAATATGGTGGTACTGAGTCCCAGAAAGACCATGCCCCCTGCAAGT

15

TCCATAGCATCTCCCTTCCACATGGATTGGGTGTTGAGTCAGTCAGCAAAACCTCATGACTTGAAAGTAA

GTGCCATTAA

SEQ ID NO:69

>Corn 1AW017984;

VPRSWIWLPLRRTVYFGTSVPTIFRGLTTEEIQQCFWRGFVCVRGFDRTVRAPRPLYARL
20 HFPASKVVRGK

SEQ ID NO:70

>Corn 1AW017984;

CCTGAAACAACTCAAATAACGGCCGATGAGGTTACATTGTTATAGTATATGATCAAAGAA

25

CATGTATGACCATTGTACAATAGGCCATCTTCATGTTGTCACAAACACTGATTCCACTC
ACAGGTACAATTAGAAACTAGACTATGCACTGTGTTTTTTCTGCTATATTGAATT
CAGAGTAAGAAATATGCACTCACAGAACATCTAAAGCTGTTACCATAGATGAAACATAA
TTAACAGGTATGTCGCCTCAAAATTGGGTTAGAACTTGGAGGTGATTGGCGACTTA
TTGGGTCACTCTCATTTAGGGCTATCCATCCGTTTGTAGCTCCTCTATATTCTTC

30

GATGTACCTATTATCTTCGACGCTTGGCGCTCCAGGCTTTTGCCCTAAACAAACCTTC
TGGCAGGAAAATGCAACCTTGCAAAAGGGGCTTGGTGCCCTTACTGTCCTATCAAAGC
CCCTCACACAAACGAATCCTCTCAGAACAGTGTGTTATCTCTCAGTCGTTAAACCTC
TAAATATTGAGGAACAGGTTCCAAAGTAAACAGTTCGTTGGGAGGTCCCATATCC
AACTTCGTTGGGAC

35